// NO.24-076 | 11/2024

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

November 8, 2024

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 analysis finds 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 and Zürich. We are thankful to Marco Bertoni, Ellen Greaves, Antonio Miralles, Olmo Silva, Bertan Turhan, Ulrich Zierahn-Weilage, and Elena Fumagalli 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, 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 opponents argue that tracking disproportionately harms students assigned to lower tracks, and only benefits students assigned to higher tracks, thereby exacerbating educational inequality (Reichelt and Eberl, 2019). However, causal evidence on the effects of tracking itself is mixed; some studies find a positive effect (see e.g. Carrell and Kuka, 2018; 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; Abdulkadiroğlu 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 observable choice sets. Thus far, however, whether tracking acts as a selective filter for ability or confers a separate benefit on academic achievement is inconclusive.

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 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 (Abdulkadiroğlu 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 by 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 0.08 standard deviations for mathematics and reading, respectively. The university aspirations of 10th grade students also benefit from an increase of approximately 0.08 standard deviations. 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 across almost the full distribution of prior academic attainment, not only for students at a specific threshold or universal cutoff. 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 females and students from low SES 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 average baseline achievement differs between the highest and the intermediate track by 0.17 standard deviations. Conditional on the leave-own-out average of the baseline achievement distribution, we adopt the same method used in the main analysis and estimate the effects of attending higher and lower peer-quality program for inframarginal students who are always assigned to the highest track. We find that attending a higher peer-quality program has a weakly significant positive effect on average standardized achievement, but via the channel of reading test scores for boys. Peer effects in terms of academic ability are therefore unlikely to be driving the main results (which predominately operate via the channel of mathematics test scores). We do find evidence that peer behavior is important, however, particularly for girls mathematics scores (though again, this is unlikely to be driving the main results).

This study relates to a large body of literature focused on the affects 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, attending out-of-district, etc.), and unlike previous studies in the literature (e.g., Carrell and Kuka, 2018; Dobbie and Fryer, 2014; Ding and Lehrer, 2007), we demonstrate that the positive results we obtain are likely not driven solely by peer ability. 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 is centralized—we are able to focus on the distributional consequences, a matter that has received comparatively less attention.

Evidence on tracking thus far 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 Matthewes, 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 identi-

[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 admitted students to the highest track.

fication to students at a specific cut-off, and is thus potentially uninformative about the effects for inframarginal students further away from the admissions threshold.

Although we also take an RDD-type approach in this paper, one of our key contributions is that, unlike previous studies that are only able to compute effects at a specific margin, we are able to study the effects of high track attendance across a wide range of the achievement distribution. The context of the Hungarian system, with many small programs in different "markets", gives us greater variation in cutoffs and therefore allows us to use the admission criteria of individual small programs as localized cutoffs. We find that while students with lower prior performance on standardized tests are less likely to be admitted to the highest track, on average, students from the lower tail of the prior achievement distribution also benefit 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 can 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 individual 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 results, and Section 6 concludes.

2 Institutional Context

In Hungary, educational tracks are hierarchically ordered such that your choice of educational pathway from secondary school onward is determinate of future education and employment opportunities. Most students apply for their preferred secondary education program in the 8th grade.[3] Students apply for specific school-program combinations and must decide between three tracks: vocational training schools, vocational secondary schools, and grammar schools. Grammar schools comprise the highest track in which students follow an academically oriented curriculum in preparation for the maturity exam in 12th grade. The intermediate track—secondary vocational schools—are mixed programs that combine academic study with vocational training after the 10th grade, though students can opt to take the maturity examination. In the lowest track—vocational training schools—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.

The structure and progression of the Hungarian system are outlined in detail in Figure 1. In this paper we focus on students at the margin of admission to the highest track—grammar schools—and estimate the value added of attendance on student test scores and university aspirations two years post-assignment. Students attending grammar schools 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 secondary vocational schools, on the other hand, students are introduced to a vocation-oriented area of study in 11th and 12th grade. 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 which leads to a professional qualification (*"szakmai vizsga"*).

Better preparation at the secondary school stage can therefore significantly affect later education and career opportunities. 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.

When choosing school-course combinations, each student is allowed to submit a

[3] A small number of high track programs offer enrollment after 6th grade. These programs do not participate in the 8th centralized assignment procedure. They are not a concern in this setting, given we focus on students *at risk* of highest track assignment, not always seated students.

strict preference ranking of arbitrary length, and choices are not geographically restricted. Since the year 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 the capacities of schools in Hungary far exceed the number of students due to demographic change, schools are typically able to offer dormitory places to most applicants. Programs may base their admittance criteria determining how applicants are ranked on a number of factors. These include results from centralized examinations organized at the beginning of 8th grade, 8th grade in-school achievement grades, and some particularly popular or selective programs also request students participate in oral interviews. Some programs rely on all of these criteria, while others rank applicants solely based on in-school performance during the 8th grade. Additionally, some programs may prioritize students with a particular religious affiliation or student-specific characteristics, such as an enrolled sibling.

Typically, programs rank their applicants according to a weighted average of the grades at primary school, entrance exam scores, and interview scores. The weights are determined by the programs, though they are subject to some constraints. Based on these rankings and a list of student preferences, the final assignment is then organized at the national level via a centralized assignment mechanism. The matching is performed by computer software using the student-proposing deferred acceptance (DA) algorithm (Biró, 2008), described in greater detail in Appendix B. The Hungarian mechanism does not use a randomized lottery-type tie-breaker. Schools must generate strict rankings over students, though they retain the right to not 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 a combination of individual-level data on individual students and their parental background, administrative data regarding the matching procedure, and school-specific information for the school-course combinations.

KIFIR.[4] This is an administrative dataset containing the outcomes of the national centralized matching scheme in Hungary which covers the universe of 8th grade students in Hungary applying for secondary education. We have access to the 2015 outcomes. The data contains information on the students' rankings over school-course combinations, the school's ranking over the students, and the final matches. 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 are matched to a school in the first round.

National Assessment of Basic Competencies (NABC). 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. The tests are not designed to measure student performance according to a specific curriculum, but rather fundamental competencies. 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. Currently, we have access to test results for grades 8 and 10 from 2015 and 2017, respectively. In addition to raw 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 GPAs, both overall and by subject), classes repeated, family background (e.g. family composition, parental education, parental occupation, and employment status), the student's career aspirations, subject preferences, and any extra-curricular activities. Furthermore, the data set includes information from school questionnaires that provide, amongst other factors, information on the school, school site, and school size.

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

[4]The abbreviation KIFIR can be translated as "secondary enrolment information system".

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]

4 Empirical Strategy and Identification

The fundamental problem when estimating the causal effect of school-track attendance on outcomes several years post-assignment is the non-random sorting of students. If certain schools attract better students, higher test scores post-assignment cannot be solely attributed to the school. Under centralized assignment, offers made at specific schools are determined by student preferences over schools and the rankings by schools over students. These preferences and rankings are thus 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 attendance at specific schools.

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 a student's preferences over schools and their local risk of assignment at each of their preferred schools. This is more practical than fully conditioning on student preferences when the number of preference "types" is very large. Second, by only comparing applicants in a narrow window around school-specific cutoffs, similar to a regression-discontinuity design (RDD), we eliminate bias arising from correlation between a student's potential outcomes and their relative rank position. A brief description of the underlying theory and the resulting estimation strategy are discussed in the following, though in-depth discussion of most issues can be

[5] Table A.1 in the online appendix tests for selective attrition to ensure the response rate of missing NABC observations in the 10th grade is not a function of their high track assignment probability, previous student performance measures, or individual student characteristics.

found in Abdulkadiroğlu 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 non-lottery 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 θ_i denote student i's list of preferences (henceforth, student type). Let student i's ability be denoted by ϵ_i. In this abstraction, students are matched to schools by a DA algorithm that takes only student type (θ) 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 value one when student i is assigned to school j. $Z_i = (Z_{i1}, \dots, Z_{iJ})$ 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 θ_i.

As noted in Abdulkadiroğlu et al. (2017), however, in practice, there are typically nearly as many preference combinations as there are students. Consequently, full type conditioning is often not feasible or leaves very few degrees of freedom for empirical analysis. Instead, Abdulkadiroğlu 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*, where for student i assignment to school j is defined:

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

Abdulkadiroğlu 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:* $\epsilon_i \perp\!\!\!\perp Z_i | p_i$.

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-dimension propensity scores thus leaves far more degrees of freedom than full type conditioning.

Abdulkadıroğlu et al. (2022) generalise 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 realisations close to key cutoffs, 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 cutoffs* denoted τ_s for school s. For any school s, τ_s is determined by the tie-breaker of the last student (highest tie-breaker value) assigned to s. For each school-specific tie-breaker cutoff, we define an interval $(\tau_s - \delta, \tau_s + \delta]$ where the parameter δ is a bandwidth analogous to that used for non-parametric RDD estimation. In the limit, as δ shrinks to zero, the probability that a student has a tie-breaker value that clears cutoff τ_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 cutoff. This probability is given by $0.5^{\hat{m}_s(\theta, R_i)}$ (line 3). Finally, line 4 gives the assignment probability at school s where student i does not surely qualify, which is the product of the disqualification rate at the applicants' preferred schools (given in line 3) and the qualification rate at s, where the latter is given by 0.5.

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 β, 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 δ shrinks to zero), D_i is independent of potential outcomes conditional on an estimate of the local high track propensity score.[6] These

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

results suggest a 2SLS procedure with second- and first-stage equations that can be written in stylized form as:

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

where β 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 cutoffs determined by the chosen bandwidth δ.

These local linear controls serve to control for imbalances between students just above and below cutoffs 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 analogous to (5). The sample used to estimate (4) is limited to applicants with high track assignment risk, thus $g_1(Ri)$ is defined analogous to $g_2(Ri)$.

Note that saturated regression conditioning on the local propensity score eliminates applicants with estimated score values of zero or 1 because track assignment is constant for these applicants (e.g., all applicants with a score of 1 are assigned to a high track program). We therefore say an applicant has *high track assignment risk* when $\hat{\psi}_G(\theta_i, R_i, \delta) \in (0, 1)$. Given the small size of programs in terms of number of seats, we do not compute bandwidths separately for each program but choose one bandwidth. When there are fewer than two in-bandwidth observations on one or the other side of the relevant cutoff, the bandwidth for that program is set to zero.

One benefit of our approach is that by allowing a flexible definition of the admissions cutoff, which varies by program, we can study the effects of high track attendance across the full achievement distribution. This allows us to overcome a typical limitation of an RDD-type approach, wherein estimates are localized around a universal admissions cutoff and the Local Average Treatment Effect (LATE) is estimated only for marginally accepted/rejected students. When cutoffs are globally applicable, it is not *a priori* clear whether applicants not in this empirical bandwidth around the cutoff would also benefit from high track attendance. In our setting, however, there is a large degree of variation in school-specific admissions cutoffs for both high track programs and intermediate track program. Figure 2 provides kernel density estimates illustrating substantial common support between our at-risk sample and almost the full spectrum of prior achievement. This allows us to use the admission criteria of individual small programs as localized cutoffs, and estimate the LATE for different points on the baseline achievement distribution.

4.2 Validation of empirical design

The empirical strategy uses students with non-degenerate assignment risk to approximate randomized high track enrollment. This requires that, conditional 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 in the Hungarian setting.

First, Table 1 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 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 2 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 2, 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 empirical bandwidth ($\delta_N = 0.25$). The bandwidth is selected in order to retain sufficient observations, while also being sufficiently narrow to balance the characteristics of students with the same local propensity scores who hold offers for different tracks (i.e. high track programs vs. intermediate track programs). 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 effect of selection and omitted variable bias on estimates.

5 Results

As described in Section 4, we instrument enrollment in the highest track with an offer of a place in a high track program. Table 3 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 3, the first stage coefficient suggests that an offer from a high track program increases the likelihood of high track enrollment by 66 percentage points.

Our main 2SLS results are reported in Table 4, and show 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. On average, attending the highest track leads to an improvement in test scores of approximately 0.11 standard deviations in the third fully-saturated specification. When disaggregating this effect by subject-specific standardized test grades, 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. University aspirations are also positively affected by attending the highest track, at .08 standard deviations.

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 4 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 4 are computed based on a sample which 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. First, Figure 3 shows that the main results are not driven by 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 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 4 also suggests that students with low baseline achievement in particular may 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. Finally, Figure 5 suggests that effects on university aspirations are similar conditional on parental higher education attendance, though there is some suggestive evidence that females, students with low baseline academic achievement, and low SES benefit more.

Any differences in effect size by, e.g., prior achievement, do not *necessarily* imply that lower ability students benefit more from attending the highest track because of some feature of the programs themselves. Rather, this could potentially be mechanistically explained by a larger quality differential between high track programs and the student's alternative non-high track assignment at different points on the prior achievement distribution. That is, the backup school for a bottom tercile student who "loses" the assignment lottery could be statistically worse than the backup school of a top tercile student. To rule this out, we explicitly test for school quality differences conditional on baseline achievement, among other factors. To do this, we construct counterfactual assignments for each student in the selected sample.

Students are alternatively assigned to the “nearest” alternative program (in absolute terms) to their actual matched program based on student preference rank, and conditional on: i. the student being deemed acceptable by the alternative program, and ii. the alternative program being a non-high track program if they are matched to the highest track, and vice-versa. We then use the counterfactual class-average leave-own-out baseline test scores as a proxy for program quality. Figures 6 and 7 illustrate, for mathematics and reading respectively, the distribution of differences in program quality between an individual’s matched program and their counterfactual assignment by terciled 8th grade baseline achievement. On average, there does not appear to be 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 8 shows no notable divergence in the distributions of differences in program quality for matched vs. counterfactual assignments by SES, gender, or settlement type (i.e., it is unlikely that in more rural areas, the quality of the next alternative program should a student “lose” the assignment lottery is significantly worse than in urban areas).

Overall, heterogeneity analyses reveal estimates of the main effects that are broadly similar. This is consistent with earlier literature that suggests students not admitted to the highest track are disproportionately harmed by tracking, given the differences in track accessions by socioeconomic background characteristics (e.g., Reichelt and Eberl, 2019).

5.2 Mechanisms

In Table 5 we examine potential moderating factors, in order 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 and, if anything, girls 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 SES, on average. Further, column (1) results suggest that peer quality in terms of baseline achievement scores differs between high track programs and intermediate track programs by 0.17 standard deviations.

This peer achievement gap is also evident when examining the distribution of differences between actual and counterfactual assignments (e.g., in Figures 6 and 7), wherein school quality is proxied with class-level matched-student average baseline test scores. Given that evidence from the prior literature suggests peer effects may lead to educational spillovers (see, e.g., Carrell and Kuka, 2018; Dobbie and Fryer, 2014; Ding and Lehrer, 2007), in part the results obtained thus far may be driven by classroom or cohort-composition effects. Peer effects can be characterized as educational spillovers, 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 effects.

The specific channel we test here is whether students assigned to the highest track benefit more from the presence of so-called "high-quality peers". We, therefore, test whether peer composition has a causal effect on student 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 while restricting the sample only to those students who are sure to attend the highest track conditional on an empirical bandwidth ($\delta_N = 0.5$). 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 programs-specific admissions cutoffs for high track programs.

Table 6 reports the results of these analyses with fully saturated propensity score controls. Estimates from the fully specified model in column (3) show that attending a school with higher-quality peers slightly improves test score outcomes in a causal sense, though the effect is not particularly significant. In Table 7 we disaggregate the effects reported in Panel B of Table 6 by subject, and repeat the analysis while distinguishing between males and females. We find some suggestive evidence that males benefit from high-quality peers in terms of reading scores, though overall peer effects are small and not particularly statistically significant. It is therefore unlikely that the effect of high track attendance on student outcomes estimated at the margin of high track admission in the main results is driven by peer quality effects in terms of academic ability, given the main effects predominately operate via the channel of mathematics test scores.

We then test two alternative channels of peer effects given a more recent literature suggests peer effort and behavior matters at the classroom level (see, e.g., Bietenbeck, 2020, Dong and Yu, 2023). Using pre-assignment behavior grades recorded in the 8th grade, and 8th grade diligence scores 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 an empirical bandwidth ($\delta_N = 0.5$). Tables 8 and 9 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 10 and 11 report those for above-median peer diligence scores. The results suggest that not only are peer behavior and diligence important, but in the fully saturated models (column 3) peer behavior effects are particularly important for females' mathematics scores.

To a lesser extent peer behavior is important for both male and female reading scores, though the effects are both smaller in magnitude and less statistically significant. Given the heterogeneity analysis conducted as part of the main results reveals the effect of attending the highest track is slightly larger for males than females, the positive effect of peer behavior on female mathematics test scores is unlikely to be driving the main results. It is nevertheless an important finding, given it on the one hand 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 males and females.

Overall, we find only limited evidence that the main results are driven by the presence of higher-quality peers 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 peer effects 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 and diligence do appear to be generally important.

6 Discussion

In this paper, we use a unique institutional framework and novel identification strategy to show that high track attendance substantially improves student performance two years post-enrollment; by the time students are in 10th grade, functional measures of competence in reading, mathematics, and other key skills are notably higher for students attending a high track program relative to those attending the intermediate track. However, unlike previous works in the literature that ascribe this value-added to peer effects, we show that—for those always seated at high track programs—there is only limited evidence peer effects in terms of ability affect average 10th grade test scores in a causal sense. Rather, we find evidence that peer behavior and diligence have spillover effects on own performance.

This matters on several dimensions given the institutional context of the Hungarian setting. Though we find tracking may amplify inequalities in skill acquisition—given assignment to specific tracks or school-types determines future attainment independent of prior ability, gender, or SES—there is unequal accession into these tracks. This is important, as both students attending the intermediate track and those attending the highest track are able to take the maturity exam at the end of 12th grade and subsequently access higher education. Further, the vocational dimension of schooling at secondary vocational schools has not yet commenced in 10th grade. For this two year period, schooling in both tracks is, presumably, focused on general education, and curriculum differences should be minimal if students in both tracks are being adequately prepared for the 12th grade maturity exam. The findings here therefore have important implications for policy design.

Critics of tracking argue that tracking disproportionately harms students assigned to lower tracks, thereby exacerbating educational inequality. Our results here support this interpretation given we find no evidence of the common narrative that more able students benefit most from high track attendance. Rather, we find that the effect of attending the highest track is not only limited to students with a strong academic background, but that students from the lower tail of the prior achievement distribution benefit at least as much from high track attendance. Given the results on peer effects suggest allowing lower-performing students access to the highest track is unlikely to negatively affect the performance of their more capable peers, and that even high performing students benefit from the addition of well-behaved classmates, these results taken together imply a potential re-thinking of assignment priorities: Expanding access to the highest track for lower achieving students could benefit those who would otherwise be assigned to the intermediate track, and would not harm those already assigned to the highest school track. Simultaneously, those who receive strong behavior grades may positively benefit their classmates even in the absence of high levels of prior achievement.

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

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

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

Abdulkadıroğlu, Atıla, Joshua D. Angrist, Yusuke Narita, and Parag Pathak. 2022. "Breaking Ties: Regression Discontinuity Design Meets Market Design." *Econometrica* 90 (1): 117–151. https://doi.org/10.3982/ECTA17125.

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. jstor.org/stable/24736029.

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, D, and M 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, By 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.

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. 10.1086/711952.

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.

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, Yinhe Liang, Xiaoqi, and Shuang Yu. 2023. "Middle-achieving students are also my peers: The impact of peer effort on academic performance." *Labour Economics* 80 : 102310. https://doi.org/10.1016/j.labeco.2022.102310.

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.

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. 10.1016/j.jpolmod.2017.11.001.

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., 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.

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. 2011. "Roel van Elk, Marc van der Steeg, Dinand Webbink." *Economics of Education Review* 30 (5): 1009–1021.

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.

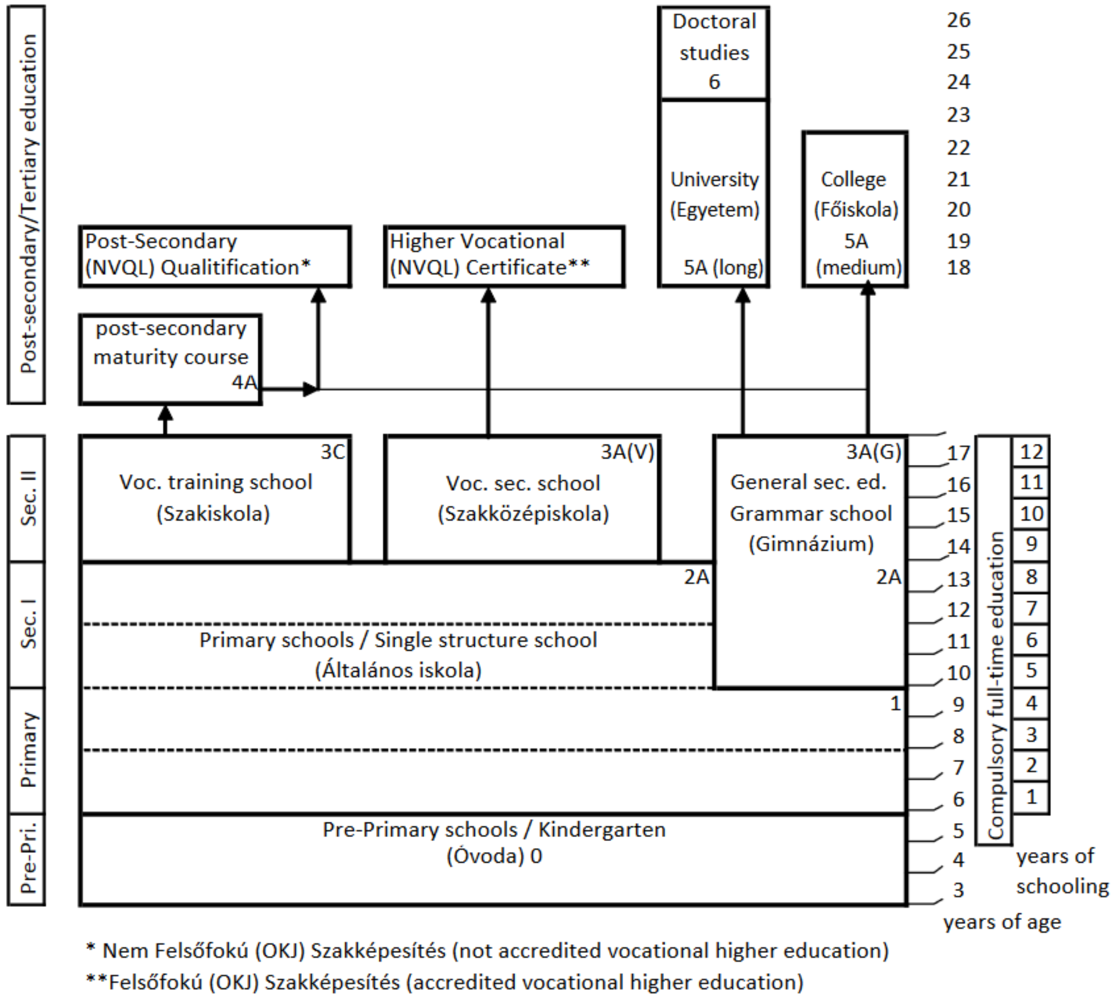
Roller, Marcus, and Daniel Steinberg. 2020. "The Distributional Effects of Early School Stratification - Non-Parametric Evidence from Germany." *European Economic Review* 125 103422.

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. 10.1007/s10797-008-9064-1.

Figures

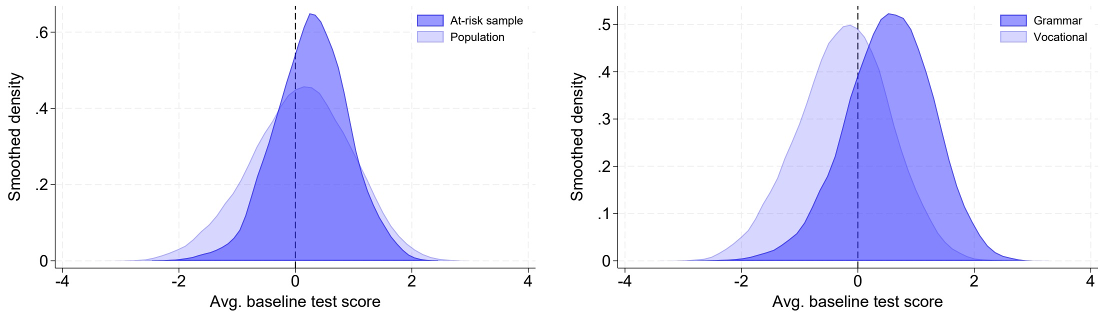
Figure 1: The Structure of the Hungarian Education System



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

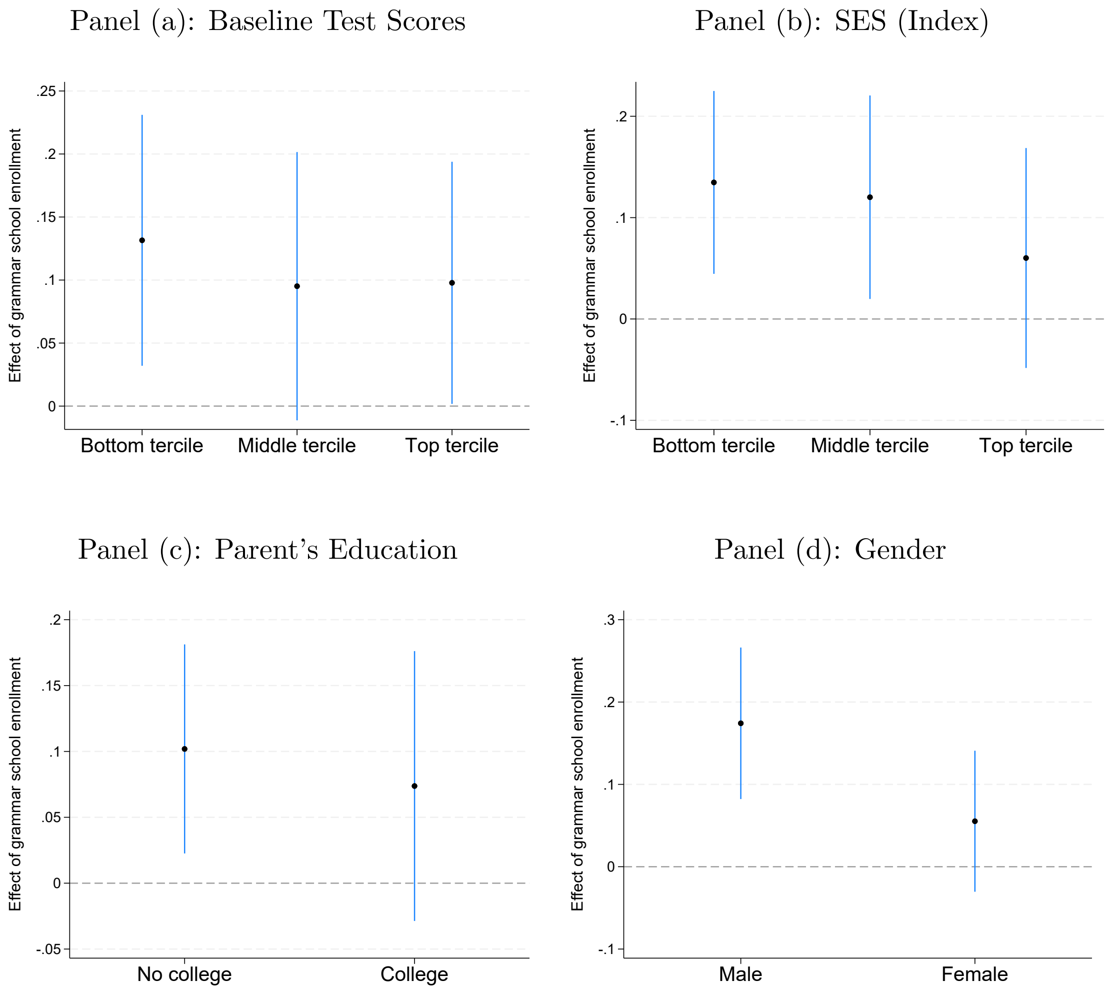
Figure 2: Distribution of Student Test Scores by Sample

Panel (a): Full and At-Risk Samples Panel (b): School Type



Notes: The figure plots kernel density estimates of average baseline test scores for (a) full and at-risk samples, and (b) for high track programs and non-high track programs in the full sample. *Source:* NABC 2015 & 2017.

Figure 3: Heterogeneous Effects on 10th Grade Average Test Scores

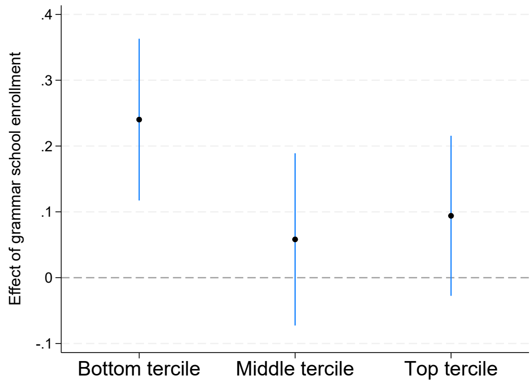


Notes: The figure plots conditional marginal effects on 10th test scores for (a) 8th grade baseline achievement terciles, (b) socioeconomic status (SES), (c) whether the student's highest educated parent attended some form of college education and (d) gender. *Source:* NABC 2015 & 2017.

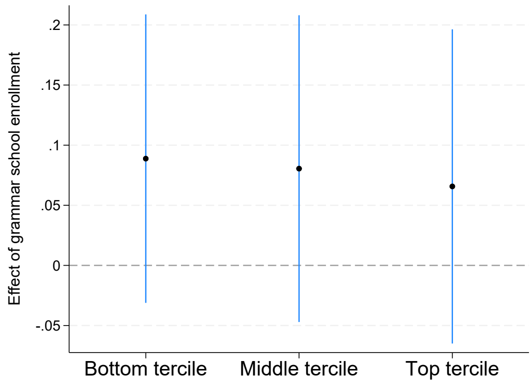
Figure 4: Heterogeneous Effects on 10th Grade Test Scores by Subject

Panel (a): Baseline Test Scores

i. Math

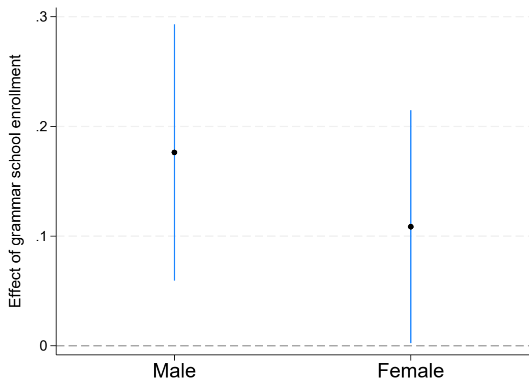


ii. Reading

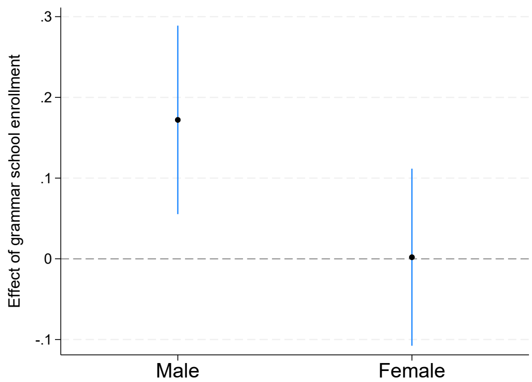


Panel (b): Gender

i. Math

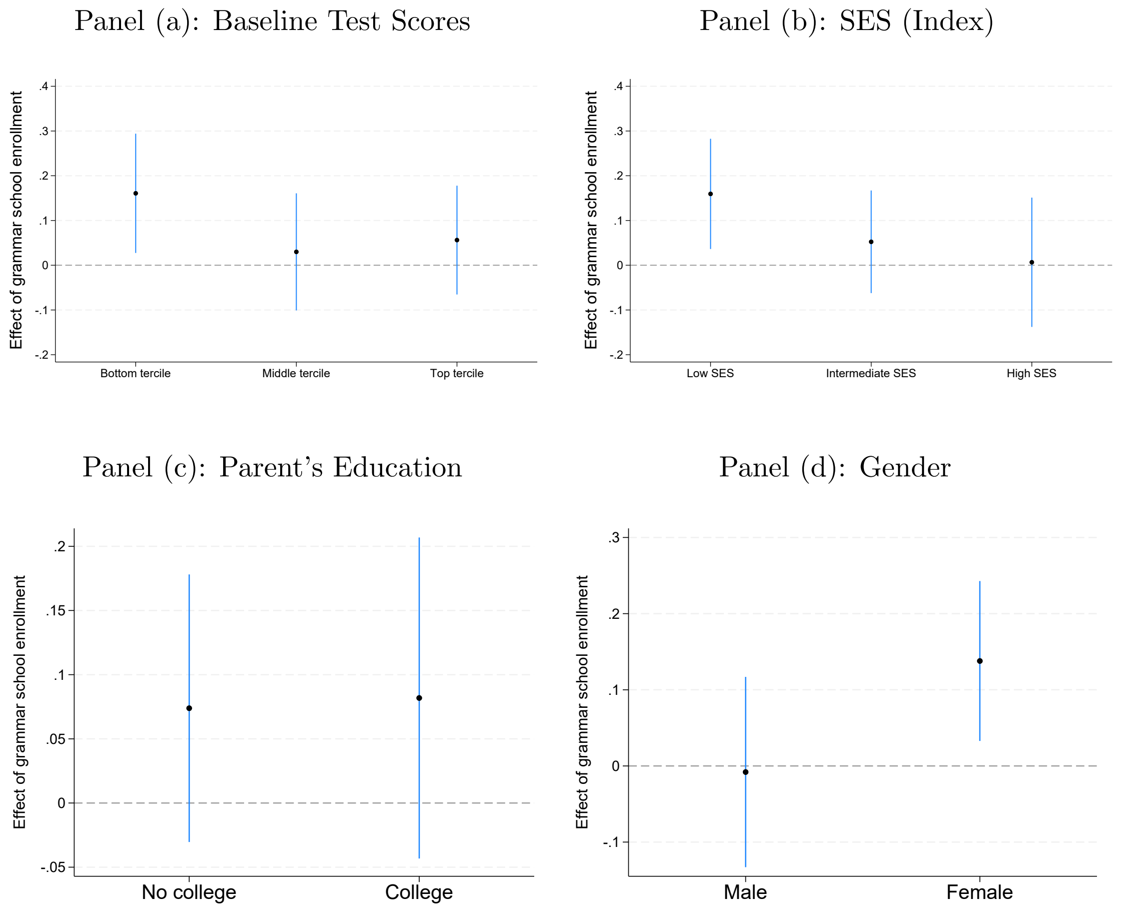


ii. Reading



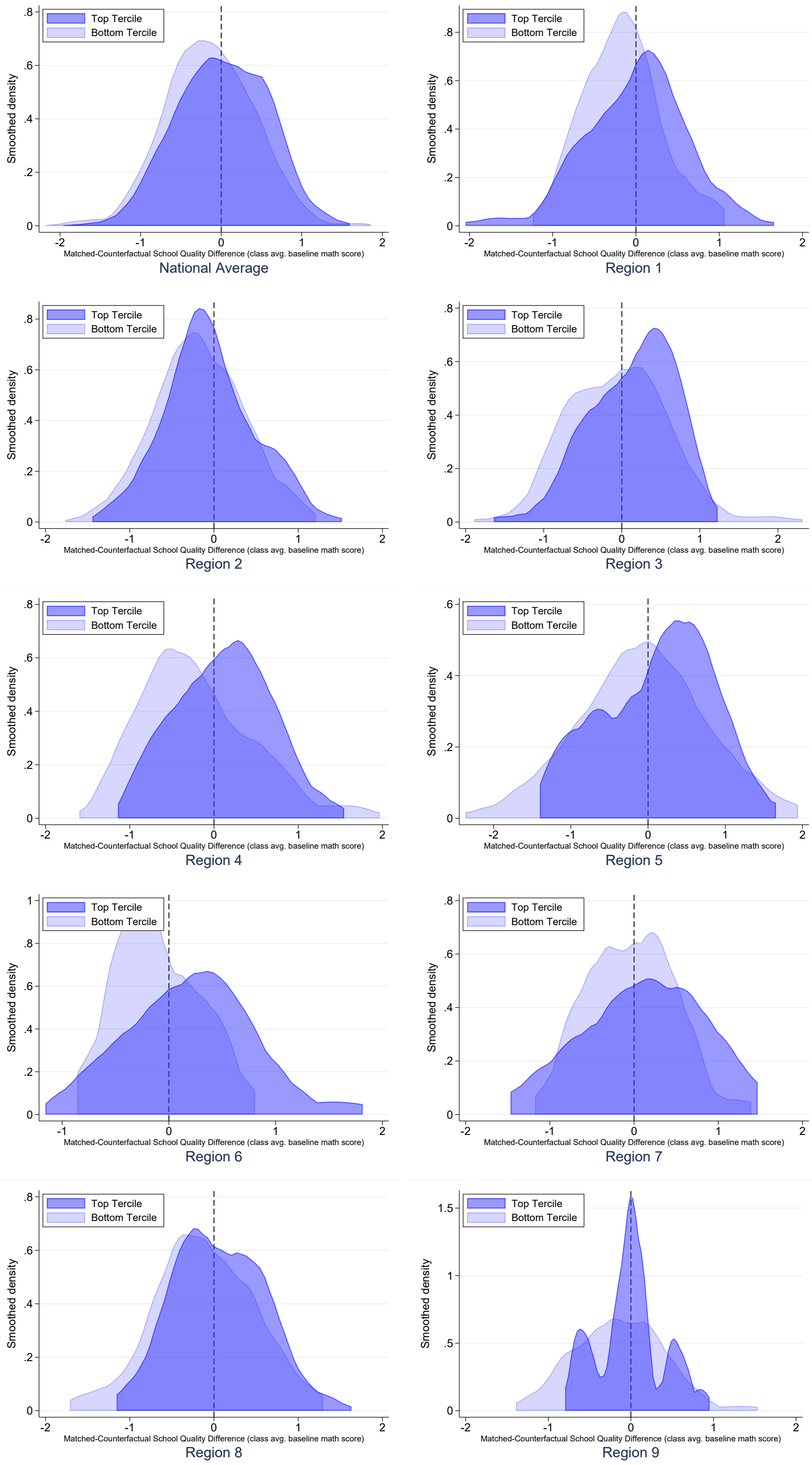
Notes: The figure plots conditional marginal effects on 10th test scores for mathematics and reading by (a) 8th grade baseline achievement terciles, and (b) gender. *Source:* NABC 2015 & 2017.

Figure 5: Heterogeneous Effects on University Aspirations in the 10th Grade



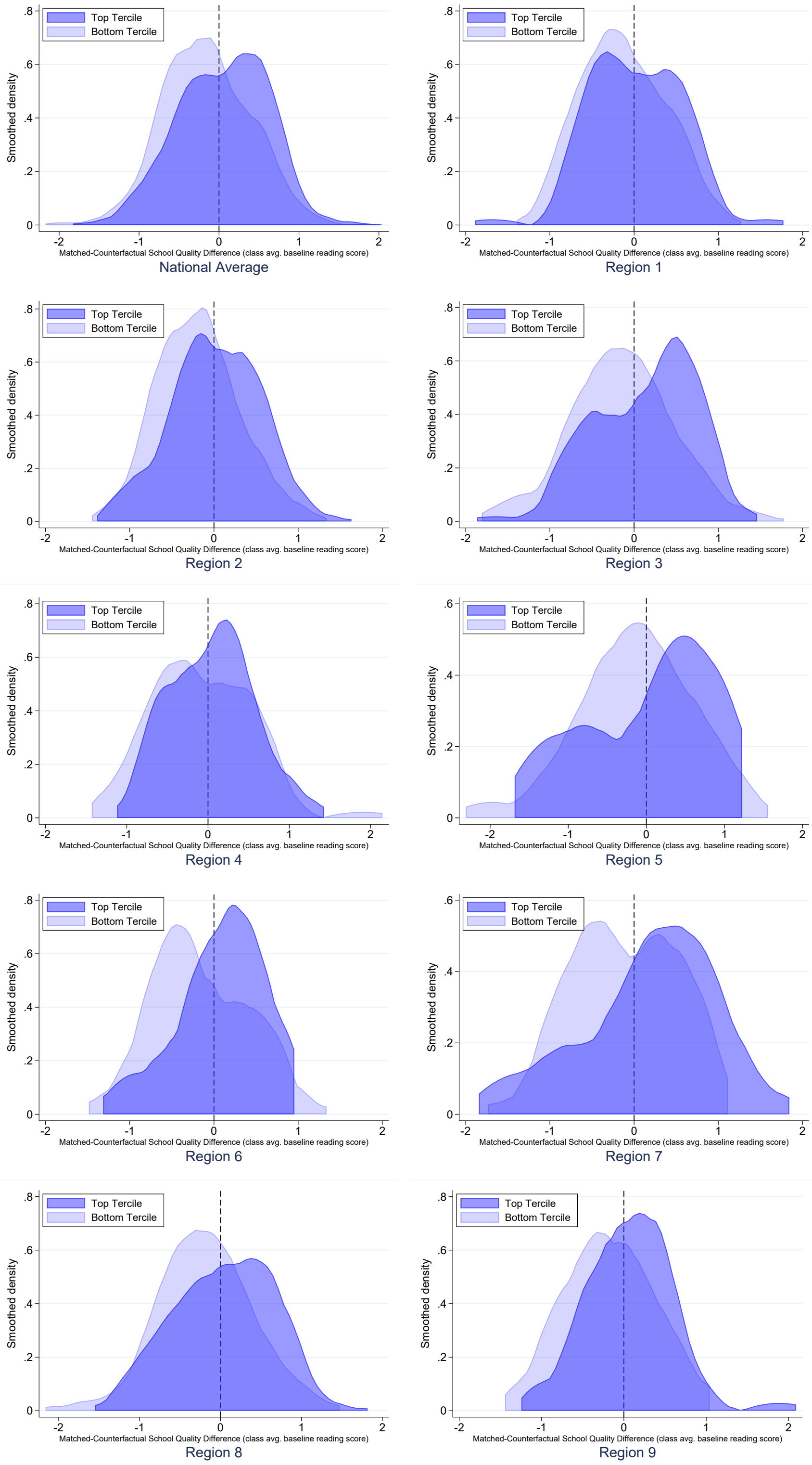
Notes: The figure plots conditional marginal effects on university aspirations for (a) 8th grade baseline achievement terciles, (b) socioeconomic status (SES), (c) whether the student's highest educated parent attended some form of college education and (d) gender. *Source:* NABC 2015 & 2017.

Figure 6: 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. *Source:* NABC 2015 & 2017.

Figure 7: Quality Differences Between Actual and Counterfactual Assignment by Terciled Baseline Test Scores (Reading)

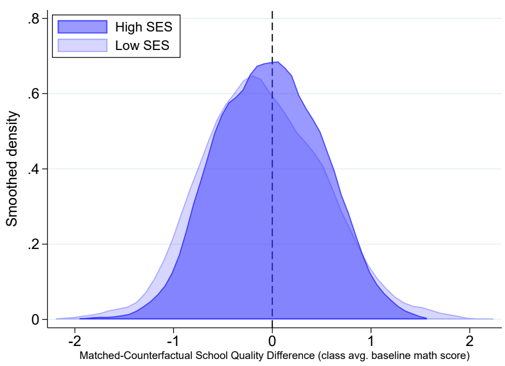


Notes: The figure plots program quality differences (proxied with class-average 8th-grade baseline achievement) between a student's matched program and their counterfactual assignment for the selected sample of at-risk students. Regions (1–9) are, respectively, Budapest, Szentendre, Hatvan, Debrecen, Szolnok, Kecskemét, Sárbogárd, Székesfehérvár, and Győr. *Source:* NABC 2015 & 2017.

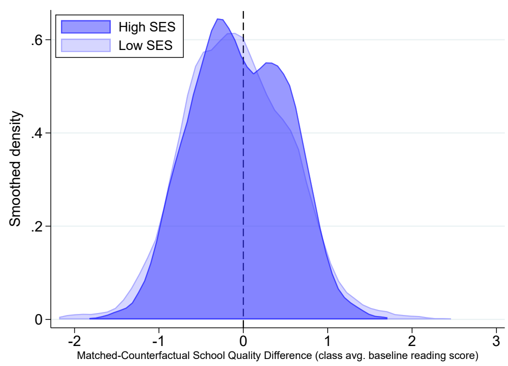
Figure 8: Quality Differences Between Actual and Counterfactual Assignment

Panel (a) SES

i. Mathematics

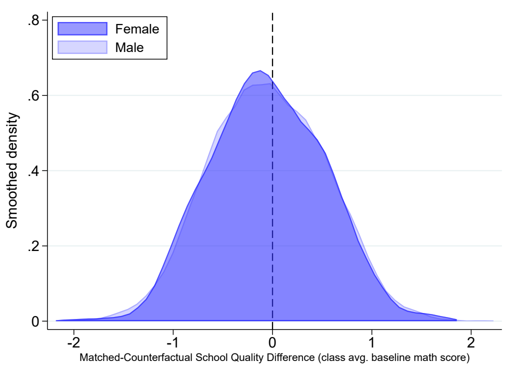


ii. Reading

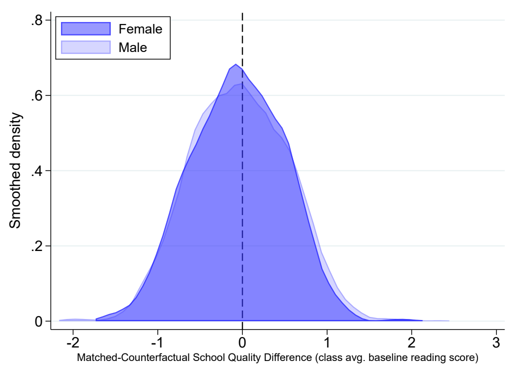


Panel (b) Gender

i. Mathematics

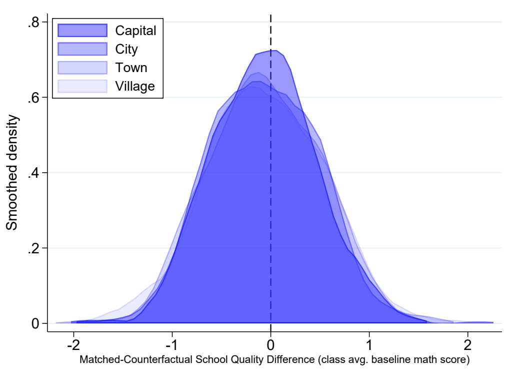


ii. Reading

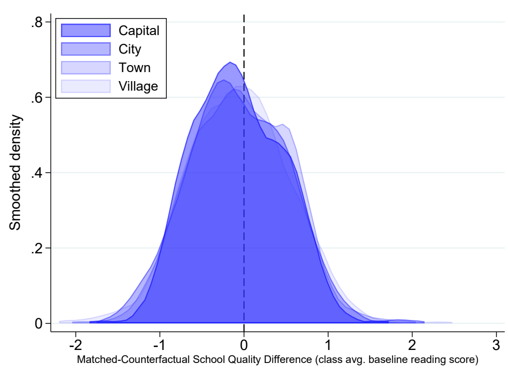


Panel (c) Settlement Type

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.*Source:* NABC 2015 & 2017.

Tables

Table 1: Sample Average Descriptive Statistics

	All Applicants (1)	At-Risk Sample (2)
Baseline standardized 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 2: Statistical Test for Balance

Dependent variable: High track offer	(1)	(2)
Baseline 8th grade math test score (std.)	0.726***	0.013
	(0.007)	(0.053)
Baseline 8th grade reading test score (std.)	0.837***	-0.028
	(0.007)	(0.055)
Female	0.157***	-0.043
	(0.004)	(0.042)
Age (in years)	-0.089***	0.028
	(0.004)	(0.034)
SES (index)	0.755***	0.050
	(0.007)	(0.060)
Social benefits	-0.125***	-0.021
	(0.004)	(0.043)
Deprived neighborhood	-0.048***	-0.015
	(0.002)	(0.023)
Single parent	-0.066***	-0.017
	(0.004)	(0.041)
At least 1 parent passed maturity exam	0.258***	0.003
	(0.004)	(0.033)
Likelihood of high track enrollment	0.230***	0.008
	(0.002)	(0.007)
Propensity score FE		✓
RDD controls		✓
N	54,631	2,518

Notes: The table presents regressions of student characteristics in grade 8 on an indicator of whether a student received a high track offer. The first column shows differences in outcomes for all students who are matched to a school in the centralized assignment. Regression estimates in column 2 control for high track propensity scores and running variables, as described in the text. The bandwidth used for column 2 is 0.25 with a uniform kernel. The sample in column 2 is limited to applicants with non-missing 8th and 10th grade test scores. Robust standard errors are in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: First-Stage Estimates: The Effect of a High Track Program Offer on High Track Enrollment

	(1)	(2)	(3)
High track offer	0.667***	0.660***	0.660***
	(0.029)	(0.029)	(0.029)
Propensity score FE	✓	✓	✓
RDD controls	✓	✓	✓
Student controls		✓	✓
Test score controls			✓
F-statistic	523.149	511.847	511.847
R^2	0.797	0.802	0.802
N	2,518	2,518	2,518

Notes: This table reports balance statistics, computed by regressing covariates on dummies indicating a high track offer. Estimates in column 3 are from models with saturated high track propensity score and running variable controls. 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 4: Effect of High Track Enrollment on 10th Grade Outcomes

	2SLS			OLS
	(1)	(2)	(3)	(4)
Panel A: Dep. Variable: Average Test Scores in 10th Grade				
High track enrollment	0.109**	0.099**	0.109***	0.181***
	(0.050)	(0.050)	(0.038)	(0.005)
R^2	0.006	0.004	0.005	0.736
N	2,518	2,518	2,518	51,135
Panel B: Dep. Variable: Math Test Scores in 10th Grade				
High track enrollment	0.164***	0.134**	0.139***	0.165***
	(0.060)	(0.059)	(0.049)	(0.006)
R^2	0.007	0.004	0.002	0.673
N	2,518	2,518	2,518	51,135
Panel C: Dep. Variable: Reading Test Scores in 10th Grade				
High track enrollment	0.054	0.064	0.079	0.197***
	(0.059)	(0.060)	(0.048)	(0.006)
R^2	0.002	0.002	0.003	0.685
N	2,518	2,518	2,518	51,135
Panel D: Dep. Variable: University Aspirations in 10th Grade				
High track enrollment	0.077	0.079	0.084*	0.256***
	(0.050)	(0.051)	(0.051)	(0.005)
R^2	0.012	0.011	0.012	0.401
N	2,030	2,030	2,030	45,740
Propensity score FE	✓	✓	✓	
RDD controls	✓	✓	✓	
Student controls		✓	✓	✓
Test score controls			✓	✓

Notes: The table presents regressions of student test scores in grade 10 on high track program attendance. Estimates in columns 1-3 are from 2SLS models where attendance is instrumented for with the receipt of a high track offer, and the inclusion of saturated high track propensity score and running variable controls. Columns 2 and 3 include student controls and baseline 8th grade test score controls, respectively. The sample is limited to applicants with non-missing baseline test scores. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: 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	0.169***	0.052***	0.202***	-0.072***	-0.105***
	(0.024)	(0.015)	(0.023)	(0.014)	(0.013)
Propensity score FE	✓	✓	✓	✓	✓
RDD controls	✓	✓	✓	✓	✓
Student controls	✓	✓	✓	✓	✓
Test score controls	✓	✓	✓	✓	✓
R^2	0.818	0.690	0.809	0.851	0.892
N	2,518	2,518	2,518	2,518	2,518

Notes: The table presents regressions of student characteristics in grade 10 on an indicator of whether a student received a high track offer. All estimates were computed with saturated high track propensity score and running variable controls, in addition to student controls and controls for baseline grade 8 test scores. Using the peer leave-own-out mean calculated at the class-level, column 1 shows differences in grade 8 baseline test scores, column 2 on gender, and column 3 on socioeconomic status (SES). Using the student's rank in grade 10 (normalized at the class-level), column 4 shows differences in grade 8 baseline test scores, and column 5 shows differences 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 6: Effects of Enrollment in a High Track Program with Higher Peer Quality (Baseline Test Scores)

	2SLS		
	(1)	(2)	(3)
Panel A: Effects on 8th grade test averages at 10th grade class-level			
High-quality peer program enrollment	0.410***	0.414***	0.414***
	(0.023)	(0.023)	(0.023)
Panel B: Effects on 10th grade test scores			
High-quality peer program enrollment	0.143**	0.132**	0.090*
	(0.064)	(0.063)	(0.049)
Panel C: Effects on 8th grade test scores			
High-quality peer program enrollment	0.075	0.070	
	(0.063)	(0.061)	
Propensity score FE	✓	✓	✓
RDD controls	✓	✓	✓
Student controls		✓	✓
Test score controls			✓
F-statistic	355	337	336
N	1,532	1,532	1,532

Notes: The table presents estimates from 2SLS models for enrollment in a high-quality peer high track program where enrollment is instrumented by receipt of an offer from a high-quality peer high track program, with saturated high-quality peer high track propensity score and running variable controls. High-quality peer high track programs are those with average 10th grade students' 8th grade test scores above the population median. Columns 2 and 3 include student controls and baseline 8th grade test scores, respectively. The dependent variable in Panel A is the leave-own-out average 10th grade class-level average of baseline 8th grade test scores. The dependent variable in Panel B is an average of a student's 10th grade test scores. The dependent variable in Panel C is an average of a student's 8th grade 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 7: Heterogeneous Effect of Enrollment in a High Track Program with Higher Peer Quality on 10th Grade Outcomes (Baseline Test Scores)

	2SLS		
	(1)	(2)	(3)
Panel A: Dep. Variable: Average Test Scores in 10th Grade			
High-quality peer program enrollment	0.143**	0.132**	0.090*
	(0.064)	(0.063)	(0.049)
High-quality peer program enrollment × boy	0.176***	0.098	0.089
	(0.066)	(0.073)	(0.056)
High-quality peer program enrollment × girl	0.108	0.163**	0.091
	(0.067)	(0.073)	(0.057)
Panel B: Dep. Variable: Math Test Scores in 10th Grade			
High-quality peer program enrollment	0.108	0.095	0.056
	(0.077)	(0.073)	(0.062)
High-quality peer program enrollment × boy	0.255***	0.046	0.039
	(0.079)	(0.089)	(0.073)
High-quality peer program enrollment × girl	-0.048	0.139*	0.071
	(0.079)	(0.084)	(0.070)
Panel C: Dep. Variable: Reading Test Scores in 10th Grade			
High-quality peer program enrollment	0.176**	0.168**	0.123**
	(0.075)	(0.074)	(0.061)
High-quality peer program enrollment × boy	0.096	0.149*	0.139**
	(0.077)	(0.086)	(0.069)
High-quality peer program enrollment × girl	0.261***	0.185**	0.109
	(0.081)	(0.087)	(0.072)
Propensity score FE	✓	✓	✓
RDD controls	✓	✓	✓
Student controls		✓	✓
Test score controls			✓
N	1,532	1,532	1,532

Notes: The table presents results of heterogeneity analyses for main results in Table 6. Each panel presents results of two models: first estimated on a pooled sample, second with gender interactions in the instrument. Columns 2 and 3 include student controls 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 8: Effects of Enrollment in a High Track Program with Higher Peer Quality (Behavior Grades)

	2SLS		
	(1)	(2)	(3)
Panel A: Effects on 8th grade test averages at 10th grade class-level			
High-quality peer program enrollment	0.115***	0.114***	0.114***
	(0.027)	(0.027)	(0.027)
Panel B: Effects on 10th grade test scores			
High-quality peer program enrollment	0.132**	0.105**	0.119***
	(0.053)	(0.053)	(0.040)
Panel C: Effects on 8th grade test scores			
High-quality peer program enrollment	-0.004	-0.022	
	(0.053)	(0.053)	
Propensity score FE	✓	✓	✓
RDD controls	✓	✓	✓
Student controls		✓	✓
Test score controls			✓
F-statistic	671	646	642
N	2,063	2,063	2,063

Notes: The table presents estimates from 2SLS models for enrollment in a high-quality peer high track program where enrollment is instrumented by receipt of an offer from a high-quality peer high track program, with saturated high-quality peer high track propensity score and running variable controls. High-quality peer high track programs are those with average 10th grade students' 8th grade behavior grades above the population median. Columns 2 and 3 include student controls and baseline 8th grade test scores, respectively. The dependent variable in Panel A is the leave-own-out average 10th grade class-level average of baseline 8th grade test scores. The dependent variable in Panel B is an average of a student's 10th grade test scores. The dependent variable in Panel C is an average of a student's 8th grade 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 9: Heterogeneous Effect of Enrollment in a High Track Program with Higher Peer Quality on 10th Grade Outcomes (Behavior Grades)

	2SLS		
	(1)	(2)	(3)
Panel A: Dep. Variable: Average Test Scores in 10th Grade			
High-quality peer program enrollment	0.132**	0.105**	0.119***
	(0.053)	(0.053)	(0.040)
High-quality peer program enrollment × boy	0.147***	0.045	0.098**
	(0.056)	(0.063)	(0.048)
High-quality peer program enrollment × girl	0.117**	0.149**	0.134***
	(0.055)	(0.059)	(0.044)
Panel B: Dep. Variable: Math Test Scores in 10th Grade			
High-quality peer program enrollment	0.169***	0.126**	0.131***
	(0.060)	(0.059)	(0.047)
High-quality peer program enrollment × boy	0.283***	0.031	0.087
	(0.065)	(0.072)	(0.056)
High-quality peer program enrollment × girl	0.057	0.196***	0.164***
	(0.062)	(0.066)	(0.051)
Panel C: Dep. Variable: Reading Test Scores in 10th Grade			
High-quality peer program enrollment	0.097	0.087	0.110**
	(0.064)	(0.064)	(0.053)
High-quality peer program enrollment × boy	0.014	0.064	0.114*
	(0.067)	(0.075)	(0.063)
High-quality peer program enrollment × girl	0.180***	0.104	0.106*
	(0.068)	(0.071)	(0.058)
Propensity score FE	✓	✓	✓
RDD controls	✓	✓	✓
Student controls		✓	✓
Test score controls			✓
N	1,532	1,532	1,532

Notes: The table presents results of heterogeneity analyses for main results in Table 8. Each panel presents results of two models: first estimated on a pooled sample, second with gender interactions in the instrument. Columns 2 and 3 include student controls 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 10: Effects of Enrollment in a High Track Program with Higher Peer Quality (Diligence Grades)

	2SLS		
	(1)	(2)	(3)
Panel A: Effects on 8th grade test averages at 10th grade class-level			
High-quality peer program enrollment	0.333***	0.337***	0.338***
	(0.034)	(0.034)	(0.033)
Panel B: Effects on 10th grade test scores			
High-quality peer program enrollment	0.131	0.128	0.120**
	(0.083)	(0.080)	(0.060)
Panel C: Effects on 8th grade test scores			
High-quality peer program enrollment	0.032	0.011	
	(0.083)	(0.081)	
Propensity score FE	✓	✓	✓
RDD controls	✓	✓	✓
Student controls		✓	✓
Test score controls			✓
F-statistic	200	195	194
N	1,127	1,127	1,127

Notes: The table presents estimates from 2SLS models for enrollment in a high-quality peer high track program where enrollment is instrumented by receipt of an offer from a high-quality peer high track program, with saturated high-quality peer high track propensity score and running variable controls. High-quality peer high track programs are those with average 10th grade students' 8th grade diligence grades above the population median as a proxy for effort and conscientiousness. Columns 2 and 3 include student controls and baseline 8th grade test scores, respectively. The dependent variable in Panel A is the leave-own-out average 10th grade class-level average of baseline 8th grade test scores. The dependent variable in Panel B is an average of a student's 10th grade test scores. The dependent variable in Panel C is an average of a student's 8th grade 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 11: Heterogeneous Effect of Enrollment in a High Track Program with Higher Peer Quality on 10th Grade Outcomes (Diligence Grades)

	2SLS		
	(1)	(2)	(3)
Panel A: Dep. Variable: Average Test Scores in 10th Grade			
High-quality peer program enrollment	0.131	0.128	0.120**
	(0.083)	(0.080)	(0.060)
High-quality peer program enrollment × boy	0.194**	0.080	0.134*
	(0.086)	(0.092)	(0.069)
High-quality peer program enrollment × girl	0.085	0.155*	0.112*
	(0.084)	(0.087)	(0.065)
Panel B: Dep. Variable: Math Test Scores in 10th Grade			
High-quality peer program enrollment	0.136	0.129	0.149*
	(0.099)	(0.091)	(0.076)
High-quality peer program enrollment × boy	0.346***	0.082	0.169*
	(0.101)	(0.108)	(0.092)
High-quality peer program enrollment × girl	-0.016	0.156	0.138*
	(0.095)	(0.098)	(0.082)
Panel C: Dep. Variable: Reading Test Scores in 10th Grade			
High-quality peer program enrollment	0.125	0.126	0.091
	(0.099)	(0.097)	(0.073)
High-quality peer program enrollment × boy	0.042	0.078	0.099
	(0.102)	(0.114)	(0.086)
High-quality peer program enrollment × girl	0.185*	0.154	0.087
	(0.102)	(0.106)	(0.082)
Propensity score FE	✓	✓	✓
RDD controls	✓	✓	✓
Student controls		✓	✓
Test score controls			✓
N	1,532	1,532	1,532

Notes: The table presents results of heterogeneity analyses for main results in Table 10. Each panel presents results of two models: first estimated on a pooled sample, second with gender interactions in the instrument. Columns 2 and 3 include student controls and baseline 8th grade test scores, respectively. The sample is limited to applicants with non-missing baseline test scores, a high track school propensity score of one, and a high-quality peer high track propensity between one and zero. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix for online publication

A Additional figures and tables

Table A.1: Testing for Selective Attrition

	(1)	(2)	(3)
High track offer	-0.027	-0.019	-0.021
	(0.027)	(0.027)	(0.027)
Propensity score FE	✓	✓	✓
RDD controls	✓	✓	✓
Student controls		✓	✓
Test score controls			✓
R^2	0.414	0.434	0.439
N	3,175	3,175	3,175

Notes: The table presents estimates from 2SLS models of high track attendance on sample attrition, where enrollment is instrumented with receipt of a high track offer. The full specification includes saturated propensity score and running variable controls in addition to student controls and baseline grade 8 test scores. The sample is limited to applicants with non-missing baseline test scores. Robust standard errors are in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B 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 should be no justified envy. The Hungarian mechanism does not use a randomized lottery-type tie-breaker. Schools are required to generate strict rankings over students. Implicitly, the "tie-breaker" at any one program is then the rank of the last admitted student.

//
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