Imola Csóka

Institute of Economics

Supervisors:

Dániel Horn, PhD

Ilona Cserháti, PhD

Corvinus University of Budapest

Doctoral School of Economics, Business and Informatics

The Effect of Elite Secondary School Programs on University Outcomes in Hungary

Doctoral thesis

Imola Csóka

Budapest, 2023

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I want to thank everyone who supported me along the journey that led to the completion of this thesis. Ilona Cserháti and András Sugár encouraged me to start my PhD studies, and Klára Major always provided me with kind and practical guidance. Working with Miklós Koren and the MicroData research group at CEU greatly contributed to developing my coding skills. It was a great motivation for me to work with Dániel Horn who helped me structure my ideas and find the way in the field of education economics. I also benefited considerably from the constructive and critical comments of the participants of two seminars in May 2022 at Corvinus Institute of Economics and CERS Institute of Economics, such as in June 2022 at JKU Linz Department of Economics. Finally, I thank my family, especially my husband for his support.

The Effect of Elite Secondary School Programs on University Outcomes in Hungary

1. Introduction

Most of the education systems in developed countries start streaming their students to different educational pathways (tracks) at the secondary school level, which usually takes place after grade 8 when students are 14-16 years old, by choosing between academically oriented and vocational programs. Within the academic track, there are special schools and programs that aim to educate the best and brightest in the cohort: elite academic high schools have competitive learning environments and highly selective admissions such as grammar schools in the UK and exam schools in the USA (see 2.1.1. International context), which is of great interest to educational researchers and policymakers. Parents bring their children to elite programs hoping to provide them with higher chances for further university education, possibly at prestigious universities. Also, to guarantee their progress with better teachers, among high-achieving peers with high social status, which could all contribute to their success not only in post-secondary education but also in their lives and labor market position later on. Although elite schools are a widely researched area in economic literature, there are still gaps that need further scientific scrutiny. For example, it remains unclear whether these “elite tracks” deliver on their promises and provide better education for their students contributing to the efficiency of the school system. It is also unclear whether these institutions only help conserve the status of already privileged students through selectivity, harming the equalizing role of the education system, or whether they provide actual social mobility to students coming from less affluent backgrounds. Hungary has a unique type of elite track, where selection happens in grade 4 or 6, much earlier than in most of the comparable elite programs. Accordingly, the institutional framework of these schools is related to early tracking in education, such as literature from elite schools.

Generally speaking, tracking (sometimes also referred to as sorting or ability grouping) means dividing students and their assignments to different educational settings (school type, program type, curriculum, courses, or classes) based on their ability, achievements, interests, or attitudes. In the US, for instance, students are sorted within the school based on achievement. In contrast, between-school tracking is more common in Europe as pupils are grouped into different schools with either academic or vocational focus. Advocates of tracking claim that it may improve the effectiveness of education by focusing on the distinct needs of students and allowing them to develop their skills according to their abilities and interests. On the other hand, the major issues raised by opponents concentrate on sustaining and increasing inequality. Tracking is more likely to benefit children of privileged backgrounds in societies that are more conservative in the sense of having less equitable social policies (Betts, 2011). Tracking reinforces the role of family background – especially if sorting occurs at an early age – thereby contributing to decreased equality of opportunity and intergenerational mobility. Tracking can change not only early educational outcomes, measured mainly through test scores, but also have important consequences on later outcomes such as post-secondary enrollment, employability, and earnings. Throughout its impact on educational pathways, sorting prevents some individuals from progressing to higher education (Brunello and Checchi, 2007).

The efficiency and selectivity of education are essential topics for public policy because education is one of the strongest tools to increase human capital and tackle intergenerational inequality. If we look at the PISA results, Hungary performs slightly weaker than the OECD average and shows a negative trend after 2009 in reading, mathematics, and science (OECD, 2019). Socio-economic disparities in academic performance are among the highest in Hungary compared to other OECD countries in several indicators: high difference in test scores between advantaged and disadvantaged students, low proportion of academically well-performing students from disadvantaged backgrounds, and high explanatory power of economic, social, and cultural status (ESCS) in performance. In the latter, Hungary scores the highest in the country group with 19,1% of the variance in reading explained by ESCS, whereas the average is much lower, only 12%. The share of top achievers - students with scores of 5 or 6 in at least one subject from mathematics, reading, and science - was 11,3% in the most recent round, trailing the OECD average (15,7%), indicating that improvements should be made at the upper end of the ability distribution as well. This high-level overview of the potential problems associated with tracking in the Hungarian education system is ambiguous. On the one hand, inequality of opportunity should generally be reduced in school systems, providing a valid argument against selection procedures or at least against their inequality-producing effects. On the other hand, reducing the gap in the proportion of top achievers compared to the OECD average could be seen as a reasonable argument in favour of selection procedures and more concentrated training for higher achievers.

The effect of elite schools and tracking are widely researched areas in academia. The primary issue for students and parents is whether applying to elite programs is worthwhile and what kind of gains (if any) can be made if admitted. For policy makers on the other hand, not only elite program's effectiveness can be interesting, but its possible impact on inequality as well. My research focuses on identifying the effects of elite programs. I compare students who entered the elite program in Grade 5 or 7 (treated group) to students who entered the general program in Grade 9 (control group) in the same academic high schools. I address the following questions:

1. What is the role of test scores and family background in admission to elite programs? Which is the stronger factor? How does it relate to inequality?
2. Do elite programs help increase students' test scores more than the alternative track (general programs)? Is this impact (if any) substantial or negligible?
3. To what extent do elite programs impact post-secondary education outcomes, such as enrollment to BA and MA level, obtaining a degree, type of major, and quality of university? Where is the greatest impact?
4. What are the most important channels of the mechanism? Is it through improved test scores, grade point average (GPA), aspiration, or something else?
5. Are the effects heterogeneous? Do students of different test scores or backgrounds experience different effects?

Based on the literature, it is still unclear whether attending elite programs results in higher test scores, and the findings regarding longer-term outcomes such as post-secondary education, earnings, and well-being (e.g. occupational rank and health) are also mixed. The reason for the lack of consensus in the literature is twofold: it stems from differences in education systems that result in context dependency, as well as methodological issues.

First, it stems from methodological issues. Selection to elite schools is usually not random which makes identifying their causal effects challenging. Second, differences in education systems result in context dependency. It is difficult to compare studies conducted in different countries because tracking practices vary greatly based on type, proportion, implementation, and starting age of tracks, for example. Furthermore, societies with different historical backgrounds and institutions respond to policies differently.

This study exploits rich administrative individual panel data in Hungary to study how elite program enrollment affects students' university outcomes. First, I analyze the selection mechanism, and show that not only test scores, but also family background influences admission to elite programs. Then I measure the effect of elite programs on test scores and post-secondary education, where I find positive effects. Afterwards, I look at the causal effects of elite program enrollment on higher education outcomes using two identification strategies, relying on the unconfoundedness assumption: ordinary least-squares (OLS) and propensity score matching (PSM). Both estimates show significant 3-4 percentage points effects of elite secondary school programs on university enrollment and completion rates. These differences are much lower than raw differences but are non-negligible, and the relative impact of elite secondary school programs on university completion and MA enrollment is much greater in magnitude compared to that on BA enrollment. Suggestive evidence shows that a substantial part of the enrollment effect can be attributed to improvement in school performance – test scores and GPA. I also offer further suggestive evidence on the potential mechanism through teacher/school quality. I also look at heterogeneity in the dependent variable (university enrollment) by distinguishing majors (STEM[1], arts, medical, law and governance), internationally recognized universities and publication performance of universities (university quality), and MA enrollment. Findings are inconclusive about STEM track choice, but I find significant and non-negligible positive effects on the quality of the university where students enroll and on the probability of whether students continue MA-level education as well. A detailed heterogeneity analysis is conducted on groups based on explanatory variables (e.g. gender, SES, test scores, primary school quality) shows that students of more disadvantaged groups benefit more

[1] Abbreviation for science, technology, engineering, and mathematics.

from elite programs. That being said, the effects on more privileged students are primarily significant, but smaller in magnitude.

My third identification strategy relaxes the unconfoundedness assumption. There are likely several unobserved sources of selection to elite programs that also affect higher education outcomes (such as motivation, parental support, non-cognitive skills, etc.). Thus, I apply instrumental variables (IV) estimation to test for omitted variable bias in the current estimates, and the results do not indicate the presence of severe bias. Moreover, I conducted several robustness tests to support the findings, and the main coefficients remain stable throughout different specifications.

The study aims to extend the knowledge relating to the effects of elite secondary school programs in international and Hungarian literature as well. The contribution of this research is threefold. First, this study examines and compares short-run (test scores) and long-run post-secondary outcomes together in a similar setting. Second, to the best of my knowledge, this research is the first to look at university outcomes in the context of elite programs in Hungary. Third, considering university quality as an outcome in the international literature regarding the effect of elite schools is rare, and the application of international rankings as a quality measure is unique.

Many limitations apply to my findings. As I mentioned above, elite programs in Hungary are unique in the sense that it is an early selection at two different time points (after grades 4, and 6) and typically within schools where the alternative is the general academic track (after grade 8) of the same school. As a result of context-dependency, my findings are not directly comparable to the standard elite school literature, in which selection occurs rather later, at a single point in time and between schools. Second, the unconfoundedness assumption (that I observe all variables that impact both selection to elite programs and higher education outcomes) is rather strong. Some of my results based on this assumption might be upward biased because unobserved factors (such as parental motivation, non-cognitive skills) presumably correlate positively with both the treatment and outcome variables. Third, while the IV estimation might solve some of the unobserved selection, the exclusion restriction might not hold.

The structure of this study is as follows. In the next part, I present several forms of tracking (grouping of students based on ability) at the secondary level in foreign school systems and in Hungary to provide the international context for this research, then I conclude the literature about the impact of early tracking and elite schools. Thereafter, I summarize the findings about the outcome variable of my model: measures of university success and quality of institutions. The data and the empirical strategy are discussed in Section 3. Section 4 presents the main estimation results, and Section 5 outlines the IV approach and robustness tests. Finally, Section 6 concludes and discusses potential policy implications.

2. Literature Review

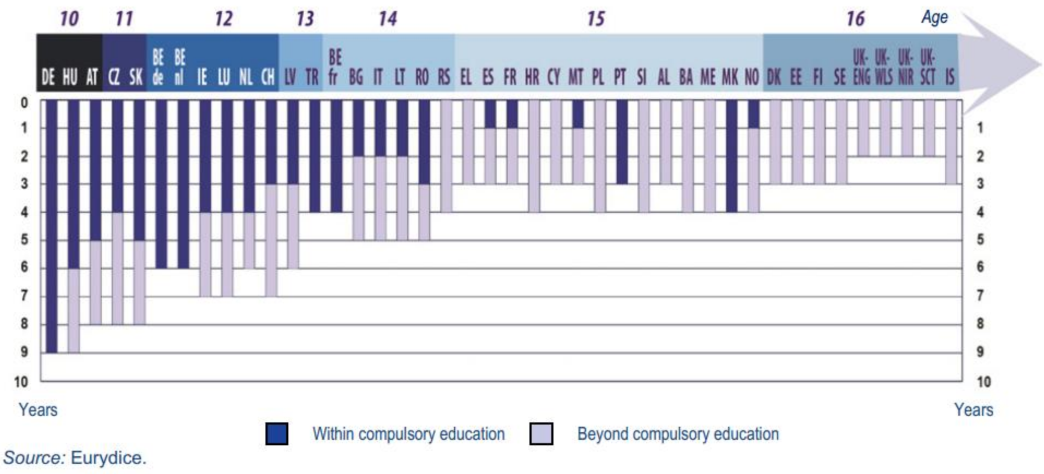
2.1. Forms of tracking and elite programs in secondary schools

Differences and lack of agreement related to the effect of tracking in the international literature stem not only from methodological issues but also from the differences in countries' education systems and tracking types. Two distinct forms of tracking exist on the secondary level: between-school tracking is more present in Europe, in contrast, within-school tracking is more prevalent in the US. In this paragraph, I compare the main types of school systems in the EU and US with a special focus on the forms of tracking at the secondary school level and describe the Hungarian education system in detail. This will provide the basis for the international embeddedness of this research. The comparison with the international findings is challenging because the elite programs in Hungary are a special type of institute which is the mixture of early tracking and elite schools. Tracking (grouping based on ability) in Hungary begins at the end of 4^{th} and 6^{th} grade only for elite programs, whereas the academic versus vocational tracking happens later at the end of 8^{th} grade. Similarly to other countries, most of the students in Hungary spend only their last 4 years of education in tracked schools, but elite selection happens earlier and those students spend 8 or 6 years in elite programs, whereas the internationally it lasts mostly 4 years. In this sense, the topic relates to the literature of early tracking, which typically examines the selection and the inequality effects (fairness) of this form of education. On the other hand, elite programs are typically separate classes within academically oriented secondary schools, which is a form of within-school tracking, so only partially comparable to between-school tracking prevalent in other countries. As a consequence, “elite programs” in Hungary are not directly comparable to the “elite schools” literature but relates to it due to the fact that high-ability students are selected there as well.

2.1.1. International context

The study of the European Commission (2020) about equity in school education in Europe summarizes different forms of tracking present in the member states. The report suggests, that assigning students to different educational pathways that have different curricula has considerable impacts on equity in education, although effects depend on how it is organized: age of first tracking, number of tracks and differentiation, size and labor-market orientation of vocational tracks, selection procedures, permeability (possibility to change track mid-studies) all matters. Figure 1 shows the starting ages of between-school type tracking: the earliest tracking age in Europe is 10 years in the case of Germany, Hungary, and Austria, whereas the majority of countries assign students to tracks earliest at age 15, and the latest by the age of 16. In the case of Hungary, the first age of selection is at age 10 for elite programs, but in general, tracking happens at age 14 when students can choose between vocational and academic orientation.

Figure 1: Starting ages of tracking and total years of schooling



Source: European Commission (2020, p.127)

The proportion of students in vocational tracks vary also widely between 15% and 75%. Differentiation is predominantly present either among vocational or general tracks, and permeability between tracks is often limited, especially from vocational to general track (although where it is allowed, schools often set conditions themselves). Usually, students are admitted to general or more demanding academic tracks through standardized exams, evaluation of the sending institute, and school entrance exams/interviews, although some countries use neither of them when assigning students to higher-level tracks. Course-by-course tracking (also called within-school tracking) is more common in secondary than in primary education, and it is present even within comprehensive school systems. Historically, most developed countries applied between- and within-school tracking as well, but during the 1970s many countries (US, UK, Sweden, Norway, and Finland) reformed their education system and either delayed the age of selection to a later age or eliminated academic/vocational streams that resulted in more comprehensive schools, decreased between- and increased within-school tracking, an example that was later followed by others such as France, Spain, and Poland (Chmielewski, 2014). In some countries, academic/vocational training and comprehensive schools coexist, but generally one is more predominant. Chmielewski (2014) compares achievement gaps across the US and 19 other developed countries between tracks in course-by-course tracking and academic/vocational streaming and finds, that course-by-course tracking is less segregated by socioeconomic status (SES) than academic/vocational streaming, yet comparable achievement gaps are present in both forms. She concludes that achievement gaps are likely attributed to some combination of track effects and selection effects.

British and American education is the most widely researched in the international literature, therefore these two systems deserve special attention. Elite programs in Hungary are comparable to exam schools in the US and grammar schools in the UK, in the sense that they both have selective admission, they are more prestigious than their comprehensive counterparts, higher-ability peers are present, these schools have academic focus and they aim to prepare their students better for university.

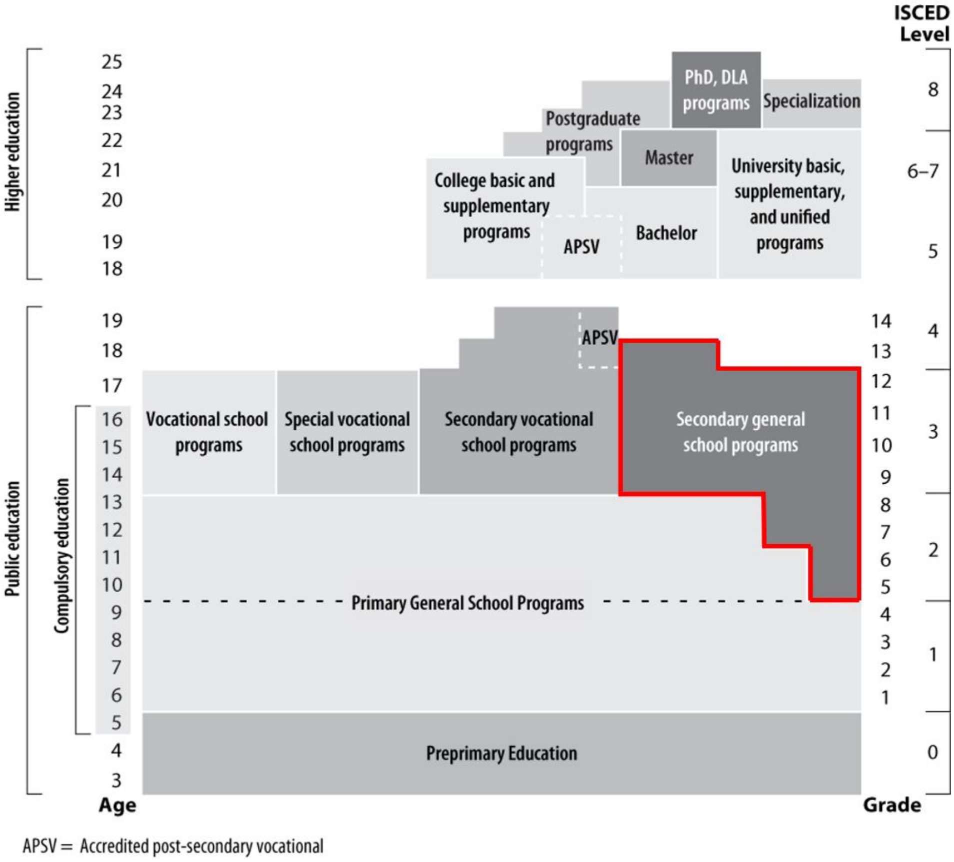
Exam schools are America’s most selective public high schools. For instance, Boston’s three exam schools (Boston Latin School, Boston Latin Academy, and John D. O’Bryant High School of Mathematics and Science) have about 2.000 students each and span grades 7-12 (admission both in 7th and 9th grade), whereas New York’s three original academic exam schools (Stuyvesant High School, Bronx High School of Science, and Brooklyn Technical High School) have about 3-4.000 students each and span grades 9-12 (Abdulkadiroglu, Angrist, and Pathak, 2014). In comparison, in one year around 7000 students enroll altogether in elite programs in more than 200 schools in Hungary, whereas around 500-1000 enroll in each of the above-mentioned US exam schools which means that size and selectivity are of different magnitude, furthermore, the timing of admission also differs: 7^{th} and 9^{th} grade is the US, and 5^{th} and 7^{th} grade in Hungary (see 2.1.2. Elite programs in the Hungarian education system). Admission is based on entrance test scores, grades, and preferences of applicants in the exam schools as well, further similarity is that these schools all provide their students with the opportunity to study together with much higher achieving peers (and also a higher proportion of white classmates) than public school alternatives. We can observe a heterogeneity between exam schools in absolute terms: the two most selective New York exam schools are among the greatest nationwide with extremely high-achieving peers (99.9^{th} percentile in SAT distribution), whereas Brooklyn Tech, Boston Latin School (99^{th} percentile) and Boston Latin Academy (80^{th} percentile) are also in the upper tail of ability-distribution, on the other hand, O'Bryant's average SAT (40^{th} percentile) cannot be considered elite in absolute terms, but still much higher than the alternative school option of those failed to enroll in exam schools, where SAT is well below the state average (5^{th} percentile) (Abdulkadiroglu, Angrist, and Pathak, 2014). Another special type of tracking in the US education system is boarding schools. These types of elite schools provide an option for students to live on its campus, which also decreases the role of parental impact on students as they live further from their families.

At the secondary school level in England, there are comprehensive schools and a few selective grammar schools where 5% of the pupils in the English state system are educated and one can get in based on academic attainment at the age of 10. The existence of grammar schools is the topic of intense political debate in England, some would like to expand it to reach higher education quality and GCSE attainment, whereas opponents argue that they should be closed due to concerns about their negative impacts on equity, counterproductivity (teaching to the test) and the well-being of students (school-related stress) (Lu, 2021).

2.1.2. Elite programs in the Hungarian education system

Various forms of tracking exist in Europe, but the Hungarian system is unique. Whilst the median student will study 8 years in general (primary and lower secondary) schools (*általános iskola*) from ages 6 to 14 and continue their studies in 4-year long tracked upper secondary school in either of the academic (*gimnázium)*, mixed *(technikum)* or vocational *(szakközépiskola*) programs, there are two 'elite' programs that stand out. Some of the secondary schools (typically with 4-year long academic programs) also offer 8- or 6-year long academic programs that cream-skim the best (highest status) students in grades 4 or 6 (ages 10 or 12) from the primary schools (see Schiltz et al, 2019), offering an early to academic secondary schools. Students are admitted to elite programs after 4^{th} and 6^{th} grade based on their GPA, individual admission interviews, and centralized mathematics and Hungarian language admission exam scores, where each school can individually decide the weights of these criteria, but the test scores should account for at least 50%. Figure 2 provides an overview of the education system in Hungary, where the analysis sample is highlighted with red. I refer to the 6-year long and 8-year long academic tracks in a secondary school together as "elite programs" (treatment) and call the regular 4-year long academic program "general". In this study, the analysis sample consists of two types of students: the treated group attends elite program (which is a subgroup of "Secondary general school programs" on Figure 2), and the control group attends general program. School type is measured in grade 10. Students in vocational tracks are not considered in this analysis, since their attributes, interests, and motivations could differ substantially from students in academic programs, and these unmeasured factors could bias the results. I identify the effect based on the comparison of two groups within academic secondary school programs (see in red rectangle called "Secondary general school programs"): difference between students of general (whose secondary education begins at grade 9) and elite programs (whose secondary education begins earlier at grade 5 or 7). I would like to emphasize, that these elite students are a selected group within the already selected students who attend academic programs.

Figure 2: The overview of the education system in Hungary

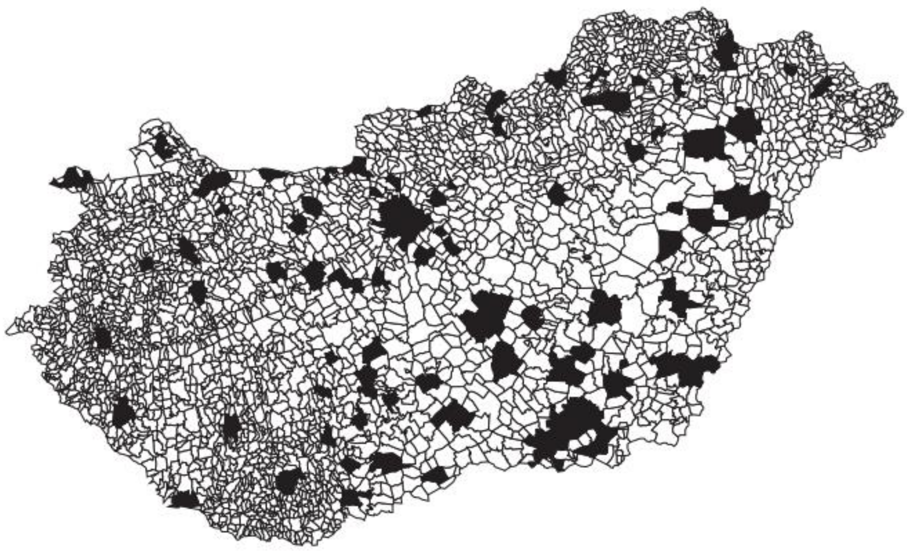


Source: Mullis et al (2016). Analysis sample highlighted in red (own edit).

These 8- or 6-year long elite academic programs explicitly offer their students a better-quality education with higher (claimed) chances of further education. Elite programs provide exactly the same qualifications (*érettségi*) as general programs but are considered more prestigious because they are longer and supposed to provide better education. On one hand, elite programs exist as the alternative to the last 2 or 4 years of education in primary schools when students are in grades 5-8, so elite students can attend secondary school earlier and stay there longer which is a case of earlier tracking. On the other hand, they are also considered as a form of ability grouping where better teachers, more advanced curriculum, and peer effects are present due to merit-based admission. The question is whether these 'elite' programs are living up to their promise and whether they are really helping to increase the proportion of students going on to university.

Elite programs are historically rooted in Hungary since they also existed during the Austro-Hungarian monarchy in the form of 8-year long academic schools that followed the 4-year elementary school. During the communist era, education was completely comprehensive and centralized, but elite programs re-appeared after the democratic transition in the early 1990s in Hungary. Although on different grounds, both liberal and conservative parties supported the concept of decreasing comprehensive education and the appearance of early selective tracks instead, the main motivation for liberals was to reach higher decentralization, for conservatives to secure the advantageous position of the higher -status people (Horn, 2010b). The establishment of elite programs halted in 1999 due to their supposed inequality effects, and the geographic coverage remained nearly unchanged since then. Although more affluent regions pushed more for the establishment of elite programs, they are geographically well dispersed with 44% of the Hungarian regions having at least one elite program by 1999, which corresponds to 6.36% of all municipalities (Schiltz et al, 2019). In Figure 4 municipalities with at least one elite program in 2017 are depicted as black, without elite programs white.

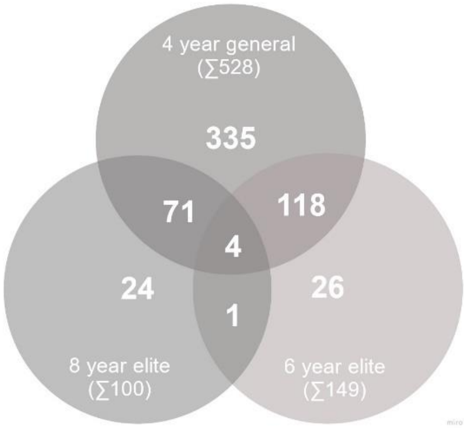
Figure 3: Municipalities with at least one elite program, 2017



Source: Schiltz et al (2019, p.80)

Vocational versus academic type of tracking exists between-schools in Hungary, although elite programs are rather within-school tracks in academically oriented secondary schools. In Hungary, 528 schools are providing academic programs, and 193 of them have elite programs as well. 100 schools have an 8-year program and there are 149 schools with a 6-year program, and in most of the cases where elite programs operate, there is a regular 4-year long track as well (calculations are based on Horn, 2013, p.41 Appendix Table A.1). In general, schools have either 8- or 6-year elite programs, there are only five schools where both early-selective track types are present. To summarize, we can say that a typical school with an elite program has a regular 4-year program as well, elite programs are typically separate classes in a secondary academic school, which has regular tracks as well (typical means around 80% here). Figure 4 shows the track type combinations within school-sites in 2010.

Figure 4: Track type combinations within school-sites, 2010.



Source: Own calculations are based on Horn (2013, p.41 Appendix Table A.1)

Around 7,5 percent of students study in these 8-year long (3% of students, it means roughly 3000 students in an academic year) or 6-year long (4,5% of students, roughly 4500) elite programs, whereas 30 percent of students study in general academic programs (Oktatási Hivatal, 2022). Statistics for 2008 stem from the secondary admission information system (*KIFIR*) showing the enrolled students in our main, baseline 6th grade cohort who completed NABC test in 2008 and got admitted for the 6-year elite program in 2008. In the two elite types around 7500 students enrolled this year altogether, and our sample is 50% resulting that the number of observations in our sample (see for example in Table 12) being 3767. Numbers change every year due to changes in the number of places provided in certain school types but remain similar in magnitude[2]. Foreign language preparatory programs also exist where they offer one additional first-year intensive foreign language training (most commonly English, sometimes German) hence the program is one year longer. This type is more common in general programs: roughly 20% of general students attend one year longer (5 years) foreign language programs, whereas this proportion is only 7% among elite students (Oktatási Hivatal, 2022 p.16).

[2] For further statistics about the secondary school system see KIFIR (Oktatási Hivatal, 2022).

2.2. Impact of early tracking and elite schools

Elite secondary school programs in Hungary are a special mix of early and elite selection, so my research relies on the international literature in the fields of "early tracking" and "elite schools" as well. While the former focuses mainly on inequalities, the latter concentrates on the impact on measurable school outcomes.

2.2.1. Early tracking and inequality

The empirical literature on the impact of early tracking has produced controversial results. In general, early tracking age harms low-ability, low socioeconomic-status students, but estimates for high-ability students are often positive, therefore tracking might contribute to the persistence of inequality (Borghans et al, 2020; Van Elk, Van der Steeg, and Webbink, 2011). On the other hand, in their experimental study in Kenya, Duflo, Dupas, and Kremer (2011) found positive effects for both high- and low-achieving groups: the direct effect of high-achieving peers is positive, whereas tracking also benefited lower-achieving pupils indirectly by allowing teachers to match instruction to students' needs.

Several studies aimed at quantifying the impact of tracking on post-secondary outcomes. These papers usually focus on higher education enrollment and completion, some also measure longer-run effects such as labor market success (employment and earnings), or other social factors such as crime, health, or teen motherhood for instance. My research is focused on university outcomes, so I concentrate on the related part of the literature and aim to collect the findings where the research question and conditions are the most similar to my setting.

Research analyzing early tracking age showed that it has an overall negative impact on higher education completion in the Dutch system (Van Elk, Van der Steeg, and Webbink, 2011; Borghans et al, 2020). They found heterogeneous effects: negative impact on lower-track and lower-ability students, however, beneficial for pupils attending the higher track and having a higher ability. Van Elk, Van der Steeg, and Webbink (2011) look at the differences in higher education completion in the Dutch environment, where tracking starts at the age of 12. They apply two different approaches in their empirical strategy: linear probability (OLS) and IV model where the relative supply of categorical schools in particular municipality types serve as an instrument for early tracking, and robust standard errors are corrected for clustering at the school level. They find a negative effect on lower track students (mavo, higher vocational) who are at the margin of the higher track (havo, higher general). OLS estimates yield significant results showing “that early tracking is associated with a reduction in the probability of completing higher education of approximately 5 percentage points” (Van Elk, Van der Steeg, and Webbink, 2011, p. 1016), even if the following variables are controlled for: age, gender, ethnicity, educational and professional level of parents, family composition, urbanization of residence and test scores at the start of secondary education. IV estimates support the OLS results, showing even larger effects in absolute value. Heterogeneity analysis reveals that girls, high-ability, high-SES students are even more negatively affected. Borghans et al (2020) investigate the effect of early tracking age on higher education completion and earning in the Netherlands. They use the relative supply of early tracking schools in the municipality as an instrument and find that the probability of obtaining a higher education degree decreases by 10 % points, earnings at age 30 by 14% on average, where the final model includes control variables on the individual level (age, gender, non-Dutch background, parental education, social class, baseline test scores of intelligence, language, math and study skills at the beginning of the secondary school) and geographic ones (urbanization and province dummies). There are significant negative effects for low- and medium-ability students, on the other hand, high-ability students presumably benefit from tracking, but the positive estimates are imprecisely estimated. The negative impact probably stems from the misallocation of students when they are sorted out early (at age 12-13 instead of 14), either assigned to a track under their potential or a too-demanding one.

Several studies focus on disadvantaged groups and intergenerational mobility in relation to tracking. In one of the earliest empirical studies in this field, Ono (2001) investigates the role of ability, social origin, and tracking in middle and high school in accessing university, analyzes mobility in the education system, and reexamines meritocracy in conjunction with institutional tracking effects in Japan. He ranks colleges into three categories based on competitiveness scores of Keisetsu Jidai, which is the Japanese equivalent of Barron’s Guide to American Colleges. He finds that social origin plays an important role in shaping the educational paths, not only advancement to college but its rank as well, while tracking reinforces the effect. Brunello and Checchi (2007) also verify the widespread belief that school tracking amplifies the influence of family history, but they find ambiguous effects: "tracking has a detrimental impact on educational attainment because it prevents some individuals from further progressing to the tertiary level of education (the diversion effect). On the other hand, the curricula offered in vocational schools seem more effective in promoting further training and adult competences (the specialization effect), thereby reducing the impact of parental background on these two outcomes" (Brunello and Checchi, 2007, p. 782). Another instance is the paper of Wanger, Dymes, and Wiggan (2017), they analyze academic tracking of Mexican immigrant students in the US and Korean immigrant students in Japan, whereas Malamud and Pop-Eleches (2011) examine school tracking and access to higher education among disadvantaged groups (from poor, rural regions, and low parental education) in Romania. Both studies find segregation effects of tracking, highlighting its social importance as well.

Terrin and Triventi (2022) look at 53 studies published between 2000 and 2021 and conclude that the mean effect size of tracking on student achievement (efficiency) is not statistically significant, although it is significantly positive on inequality. They also mention a possible "equality-efficiency trade-off" meaning that "de-tracking" might improve equality at the expense of decreasing the performance of more advantaged students.

To conclude, there seems to be consensus about the inequality effect of tracking, because its gains are concentrated among high-ability and high-SES students, and earlier tracking age harms the opportunities of students with more disadvantaged family background characteristics.

2.2.2. Elite school's impact on short- and long-term outcomes

Elite school literature is typically focused on short-term educational outcomes, and the findings are inconclusive. The identified effect sizes vary between studies, which is not surprising in the presence strong context dependency under different education systems. Most of the scientific evidence suggests that the impact of these selective schools on academic performance measured by student test scores is negligible: see for example Clarks' (2010) study of a UK district, and research conducted on US data by Dobbie and Fryer (2011) and Abdulkadiroğlu et al. (2014). Meanwhile, some studies also find positive effects: Horn (2013) in Hungary, Pop-Eleches and Urquiola (2013) in Romania, and a bit different context but rigorous experimental evidence from Duflo, Dupas, and Kremer (2011) in primary schools of Kenya.

On the other hand, there is a gap in the elite school literature about their effects on long-term outcomes. What is more, the studies in this field are generally limited to measuring only university enrollment rates. Abdulkadiroglu, Angrist, and Pathak (2014) is one of the few studies that looks at college quality as well: they measure it with selectivity as defined by Barron's, based on the competitiveness of admission. Lu (2021) distinguishes prestigious Russell Group universities, and Shi (2020) investigates STEM completion. These indicators paint a more nuanced picture and tell us more than simple enrollment rates. My study contributes to this line of research by distinguishing between field of studies, levels of studies (BA and MA enrollment) and university quality by creating a unique measure based on international rankings (see 2.3.3. Evaluation of Hungarian universities). The comparison of short-and long-term effects of elite programs in a similar setting is also rare in the international literature, so my study also aims to extend the knowledge in this area.

Elite schools have various definitions and different settings, which makes the comparison between studies challenging. Common features of these institutions are that they have a selective admission based on ability, they are more popular than the other schools, they have higher-achieving students, academic focus, and they are public high schools[3]. I conclude the relating literature based on the most relevant characteristics in Table 1, and I also collected some important additional information about the context of the studies. This summary highlights the main findings and provides a platform for comparison of short-and long-term effects.

[3] There are two exceptions: Duflo, Dupas, and Kremer (2011) study primary schools, and Horn (2013) study not exclusively high schools, but elite programs start earlier at middle school level (grade 5 or 7).

Table 1: Literature summary on elite schools

Paper	Estimated effect	Result	Design	Context	Term
Duflo, Dupas, and Kremer (2011)	ATE on test scores (18 months later)	Positive	RCT	Kenya, primary school, ability grouping	Short
Deming and his co-authors (2014)	ATE on college enrollment, quality and degree completion	Positive	RCT	USA, lottery, integration	Long
Abdulkadiroglu, Angrist, and Pathak (2014)	LATE on test scores	Zero	fuzzy RDD	USA: Boston, New York	Short
	LATE on college enrollment and quality (Barron's selectivity)	Zero	fuzzy RDD	USA: Boston, New York	Long
Pop-Eleches and Urquiola (2013)	LATE on high-stakes graduation test (12th grade)	Positive	RDD	Romania	Short
Beuermann and Jackson (2022)	LATE on test scores (10th grade)	Zero	RDD	Barbados	Short
	LATE on postsecondary school completion	Positive	RDD	Barbados	Long
Clark (2010)	ATE, LATE on test scores	Positive, inconclusive	OLS, fuzzy RDD	UK	Short
	LATE on university enrollment	Positive	Probit	UK	Long
Lu (2021)	ATE on university enrollment and quality (Russell Group)	Inconclusive	Logistic regression	UK	Long
Horn (2013)	ATE on test scores (10th grade NABC)	Positive, inconclusive	OLS, IV	Hungary (8 or 6 years long elite programs)	Short

These papers will be discussed more detailed in the following chapters, where I describe the most common econometric methods in elite school literature and present the studies by research design.

Randomized experiments

As mentioned above, selection to elite programs is usually not random, so estimating their causal effects is challenging. It is hard to find situations in which comparable students enroll in schools of different quality. The most credible causal identification comes from randomized controlled trials (RCT's), or randomized experiments where researchers use random assignment to treatment and control groups, allowing them to draw causal inferences about the effects of the interventions. By randomly assigning students to different conditions, researchers can control for potential confounding variables and better understand the causal relationships at play. Examples of this design include evaluation of tracking policies (see Duflo, Dupas, and Kremer [2011]) or admission lotteries (see Deming and his co-authors [2014]). Although, even in the case of the most reliable identification strategy, findings can be questionable due to external validity concerns. Duflo, Dupas, and Kremer (2011) apply an exceptional research design, where they study the effects of tracking in a real randomized experiment with the participation of primary schools in Kenya. Despite the high educational policy relevance of tracking, this is the very first paper to provide rigorous evidence in this field. Schools were randomly assigned to either tracked or non-tracked types, and in the latter tracking could happen either based on ability grouping or completely random assignment of students into groups. Tracked groups received one additional teacher for 18 months in grade 1 and continued in grade 2. This innovative setting allows them to identify real causal effects: they find experimental evidence of peer effects where high-achieving students increase the performance of their classmates, on the other hand, low-achieving students also benefit from tracking due to the mechanism where teachers fit curricula to their levels. Altogether, the impact was positive throughout all levels of the initial ability distribution. In the international literature, the impact on low-achieving peers - or also called the ones left behind when high-achieving peers leave the class – is often negative, although the methodology of those lacks the unique randomized setting that is present here. However, due to the unique type and form of tracking in this experimental setting, its external validity is questionable: whether the results apply in developed countries, secondary schools, and the effects are persistent in education systems on the long run as well. It would be a relevant policy insight to uncover the differences in these settings and find what barriers might prevent the realization of positive effects among lower-achieving pupils in other studies.

A few research uses lottery design in the USA. Exploiting the random assignment in the high school admission lottery in Charlotte-Mecklenburg, Deming and his co-authors (2014) identify a significant overall increase in college attainment for winners, the positive effect is more concentrated among girls. In the same setting, Deming (2011) uncovers, that adult crime and days in incarceration are also reduced among winners supposedly due to school quality and peer effects, gains are concentrated among high-risk youth. However, the context of these lottery designs where disadvantaged students are provided opportunities is not really comparable to the context of elite versus general academic setting that I examine in Hungary, where top students are selected from the good students, furthermore, no lottery is available.

Natural experiments

In social sciences RCT's are rarely feasible due to practical and ethical challenges, so natural experiments serve as the second-best solution to evaluate the causal effects of elite schools. Natural experiments in economics are real-world situations (such as a policy change or a sharp change in treatment due to a specific threshold) in which researchers can observe the effects of a particular economic event or policy change as if it were a controlled experiment, even though the event or change was not intentionally designed for research purposes. However, they also come with challenges related to identifying appropriate comparison groups, controlling for confounding factors, and ensuring that the exogenous variation truly resembles random assignment. Regression discontinuity design (RDD) is often used in the elite school literature: it compares outcomes near a threshold where a discontinuity occurs in treatment assignment due to a specific threshold value of a continuous variable. Researchers can estimate causal effects by comparing the performance of students slightly above the admission threshold to the ones who did not reach the cutoff but were close to it. For example, in their influential study, Angrist and his colleagues (Abdulkadiroglu, Angrist, and Pathak, 2014) created the concept of "elite illusion" that questioned the value-added of elite schools and showed in a fuzzy RDD setting that admission to oversubscribed exam schools at Boston and New York City have little or no causal effect on test scores and college quality. They studied three exam schools in Boston and three in New York to see whether admission to these preferred schools results in better academic achievement or higher college quality. They find at best no effect (sometimes also imprecisely estimated negative effects) on most of the outcomes, such as 7th, 8^{th,} and 10th grade math and English test scores, college selectivity defined by Barrons' (highly competitive or competitive), SAT and PSAT scores. Although test scores can be a noisy measurement of ability, results are robust throughout different schools and different measured outcomes, except for 10th grade English test scores where a little increase is detected, driven partly by gains for minorities. Peer effect and racial mix are also analyzed, and it turns out that neither the higher achieving peers nor the higher proportion of white classmates increases academic outcomes. Therefore, the reasons for parents to prefer these schools are either based on a misbelief that these more attractive schools with better peers have higher value-added, or they value these schools for characteristics other than learning outcomes, something that wasn't captured by the test scores and post-secondary variables. The authors still plan to measure earnings effects later.

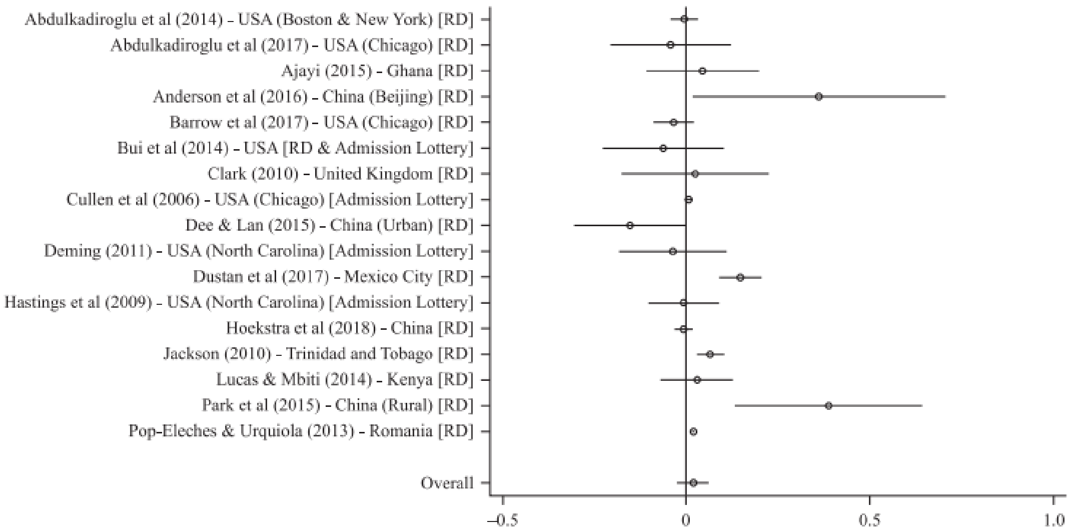
Similarly, Dobbie and Freyer (2011) found little impact on college enrollment and graduation with sharp RDD in three prominent New York City exam schools.

On the other hand, Pop-Eleches and Urquiola (2013) apply RDD in Romania and find that students who make it to the better high schools (and tracks withing schools) perform better in a high stakes graduation test (Baccalaureate exam). Admission to high schools in Romania (8^{th} to 9^{th} grade) depend on a transition score that is determined by a nationwide test score and GPA. After obtaining their transition score, students give a preference order of school and track (e.g. mathematics, literature) combinations in a nationally centralized process, which later determines the admission cutoffs. Essentially, sorting happens between schools, and also within schools (between tracks). They observe greater estimates for between-school, and smaller but still highly significant estimates for between-track cutoffs. This admission system is very similar to the one applied in Hungary, so I find the two settings quite comparable. Further, these differences in cutoffs allow researchers to explore heterogeneous effects at different points of the transition score distribution. They find that performance gains are slightly greater and more precisely estimated for cutoffs at higher grade levels, and larger in markets with fewer schools, where schools offer a more limited range of tracks (it is similar to a Budapest versus rural comparison in Hungary). The study also discovers behavioral responses: teacher sorting and changes in parental effort. Teachers with higher certification standards are more likely to work in better-ranked schools, and sorting persist even within schools from weaker to stronger classes. While average parental participation is greater in higher-ranked schools, parents reduce homework-related help if their children score just above the cutoff.

Shi (2020) studies a selective boarding school in North Carolina, where multiple admission cutoffs are present, which provides a unique opportunity to uncover heterogeneous effects with RDD. Results on the pooled sample indicate an increase in mathematics test scores and college enrollment (contrary to the findings of Abdulkadiroglu, Angrist, and Pathak, 2014), a remarkable decrease in applications for non-selective colleges but no effect on the likelihood of graduating with a STEM degree within 4 years of enrollment. In addition, heterogeneity analysis uncovers that gains are concentrated among students with lower baseline achievement, more disadvantaged backgrounds, minorities, and women, so equalizing effect is present in the boarding school.

In a systematic meta-analysis Beuermann and Jackson (2022) collect publicly available studies using quasi-random assignment to a preferred (non-charter) public school (either through lottery or selective enrollment exam) and also document that sought-after public secondary schools do not tend to improve student test scores. Although one can see on Figure 5, that there is variation in the effect sizes, which might stem from the differences between education systems, cultures and norms. The definition of the control group also impacts the results, the attributes of the alternative school placement matter a lot. Figure 5 also highlights, that context dependency is crucial when interpreting the results. Positive estimates seem to be found rather in developing countries (China, Mexico, Trinidad and Tobago, Romania), where the equity in access to elite schools might be a greater concern. On the other hand, USA and UK studies show zero effect.

Figure 5: Effects of preferred public secondary schools on test scores



Source: (Beuermann and Jackson, 2022, p. 727)

They argue that no short-run effect does not necessarily imply that long-run effects are also neglectable and explore this claim using linked administrative and survey data from Barbados. Applying RDD they find that preferred schools do not improve test scores in the short run, however, they have a positive impact on a series of long-term outcomes, such as postsecondary school completion and improved adult well-being (based on an index of educational attainment, occupational rank, earnings, and health). The novelty of this study is that “no single paper has estimated the causal impacts of the same preferred schools on test scores and also on a broad array of medium- and longer-run outcomes.” As such, it brings together the literature about short-and long-run effects of elite schools and highlights the multidimensionality of school output. Their work suggests that evaluations based on solely short-run test-score gains can be misleading regarding the welfare gains of these schools, and that parents might have additional information about the potential benefits that is not captured by the traditional variables used in the literature. Clark (2010, p.4) draws a similar general conclusion in his study in the UK: “other varieties of selective school can also improve longer-run outcomes without improving test scores”. This might explain why parents still prefer these schools despite evidence showing unconvincing effects on test scores.

In his study, Clark (2010) analyzes the effects of attending a selective grammar school in the UK, East Riding of Yorkshire, a fairly representative district. Students are assigned to these prestigious schools based on results from a primary school test, and he takes advantage of this rule by using the predicted probability of treatment (as a function of assignment scores) as an instrument for treatment and estimates the causal effects of selective schools on test results using data on these assignment test results for this specific district. The paper demonstrates that four years of selective school attendance has at best small impacts on test scores, even though students in selective schools experience substantial peer advantage. He also investigates longer-term outcomes, and find positive impact on high school course taking, and university enrollment (see detailed in the next section under “Other methods”). In their later work, Clark and Del Bono (2016) take a step further and investigate the long-run effects of elite schools including completed higher education, income, and fertility. They apply regression kink design, which also leverages discontinuities in continuous variables for causal inference. The main findings are the following: elite school attendance increases post-secondary completed full-time school attendance by 1 year on average, on the other hand, it is at best weakly related to observed labor market outcomes. Women experience positive, although imprecisely estimated impact on income, employment, and wages, but estimates for men are rather negative (imprecisely estimated), supposedly because they chose further academic education at the expense of vocational training. The impact on women's fertility is significant, elite school attendance decreases the number of children by 0,4 on average. Both papers have relied on data about cohorts of students born in the 1950s, so the implications might differ in today's setting.

In these studies, a core concept for identification was the difference at the admission cutoff. The caveat of RDD is that it only identifies effects around the admission threshold, so it does not provide any information about students further from the cut-off, which can be misleading in the presence of heterogeneous effects, especially at the right tail of the distribution of the running variable. Furthermore, they lack a comparison with similar ability students who did not apply for elite schools.

Another popular method is instrumental variables (IV) estimation. It is a technique used to address endogeneity, particularly when estimating causal relationships between variables. Endogeneity refers to the situation where the explanatory variable of interest (elite treatment in this case) is correlated with the error term in a regression equation. This can occur due to measurement error (test score can be a noisy measure of ability) or omitted variables (such as parental motivation or non-cognitive skills), leading to biased and inconsistent estimates of the causal effect[4]. An instrumental variable is a variable that is correlated with the treatment of interest (elite admission) but is not directly related to the outcome (test scores and university enrollment for instance). The role of the instrumental variable is to act as a "proxy" or "instrument" for the treatment's variation, allowing researchers to isolate its causal effect on the outcome. The key features of instrumental variables are relevance and exclusion. Relevance means, that the instrumental variable should be correlated with the endogenous variable, meaning it affects the likelihood of being "treated". Exclusion restriction is a crucial assumption: the instrumental variable should not have any direct effect on the outcome variable, its influence on the outcome should be solely through its impact on the endogenous variable. An appropriate natural experiment, such as a strictly exogenous, legitimate IV (relative supply of schools, admission test scores or geographic proximity for instance can instrument elite treatment) can yield a consistent coefficient of education on the compliers, thus can have an important role to uncover the causal effects. The caveat of IV estimation is that it shows the average effect on the subpopulation that is

[4] Other sources of endogeneity (simultaneity) are highly unlikely.

affected by the instrument so on students whose behaviour changes due to the instrument (compliers), not necessarily the whole population (effect for always takers and never takers are not accounted for) which affects the interpretation of the results. In addition, one cannot test, and almost never be fully sure that an instrument satisfies the exclusion restriction, so the careful consideration of the framework and potential sources of bias is necessary. A weak IV can be also problematic, which means that the inference is unreliable if the correlation of the instrument with the outcome variable, conditional on the controls, is too low.

Fuzzy RDD is essentially a combination of RDD and IV analysis. In fuzzy RDD there is no sharp cutoff: the treatment does not necessarily switch on at a threshold, but the probability of treatment does jump (it is discontinuous) at a threshold of a certain function (instrument). In this case, the probability of the treatment changes sharply in the narrow interval of a variable that influences the treatment. This variable serves as an instrument for the treatment. For example, Clark (2010) uses the predicted probability of treatment (as a function of assignment scores) as an instrument for treatment, whereas Abdulkadiroglu, Angrist, and Pathak (2014) applies exam school offer dummies as instruments for exam school exposure.

Other methods

Lastly, when natural experiments are not available, further econometric methods are applied that rely on the assumption of unconfoundedness (also known as conditional independence assumption or selection-on-observables). Unconfoundedness means, that all confounding variables are observed and accounted for. Confounding variables are those, that affect both the treatment (T) and the outcome (Y). In this case, selection into treatment happens only on observables, and there is no omitted variable bias, so the estimate can be interpreted as a causal relationship. Traditional regression methods that rely on this assumption include OLS, PSM, and for binary outcomes logistic regression and probit. However, drawing causal conclusions with these methods is always questionable since we cannot rule out that students of elite programs are different in some unobservable characteristics, although we can speculate about the direction of the bias if we are aware of the direction of potential correlation between the unobserved factor and the variables of interest (treatment and outcome). While these methods require stronger assumptions, they might still offer valid ways to investigate research questions when no other methods are available.

Clark (2010) found no impact of selective grammar school attendance on test scores in the UK, using RDD. He also studies longer-term outcomes with probit models and find different results. Selective schools have beneficial effects on the course taking as they increase the probability that students pursue academic courses and, more importantly, probit estimates suggest that they increase university enrollment between 2 and 5 percentage points[5], showing that they may have important long-term implications. It implies, that selective school can also improve longer-run outcomes without improving test scores.

Lu (2021) investigates whether attending academically selective schools increase higher education participation rates for students in England. Her baseline logistic regression conditions on pre-existing differences in pupils' KS2 attainment and background characteristics (such as gender, age in months, ethnicity, disadvantaged background proxies: free school meal, special educational needs, English as a second language, income deprivation), where odds ratio for attending a grammar school is 1,54, meaning that they enjoy a higher likelihood of higher education participation (54%). Effect sizes drop substantially when average student composition is accounted for proxied by a school-level variable (mean KS2 total mark in secondary school) and became 1,23, moreover, it decreases further if KS4 capped GCSE score is included resulting in an odds ratio of 0,94, which shows that grammar schools exert their effect on HE completion through improvement in test scores. She concludes that the role of grammar schools in improving their pupils' academic opportunities is mixed, since their students are associated with a higher likelihood of attending higher education institutions than their comprehensive school counterparts, however, no robust evidence is found that grammar schools improve the chances of attending Russell Group universities (more prestigious institutions), because the pattern reverses after controlling for school-level student composition.

Berkowitz and Hoekstra (2011) use admission data provided by an anonymous independent elite high school and compare the results of approved candidates who later decided not to enroll with the ones who actually enroll. They discovered that enrollment at the elite school encourages students to enroll in more selective colleges.

[5] Data is not available for girls, so probit estimation was made among men.

To sum up, effect sizes and significance for academic outcomes are mixed. RDD is a commonly used quasi-experimental setting, that focuses on the marginal student around the admission score, but its findings cannot be generalized due to heterogeneous effects throughout the ability distribution. OLS and logit models might suffer from omitted variables bias, hence they can be complemented with IV estimation to draw causal implications. Furthermore, the absence of short-term test score effects might reverse to other beneficial outcomes in the long run (such as higher university enrollment and quality) and parents take into account both while shaping their school preferences.

2.2.3. Effect of elite programs in Hungary: inequality, test scores, wages

Hungary provides a special setting (see: 2.1.2. Elite programs in the Hungarian education system) to analyze the effect of elite programs. The existing studies often focus on the inequality of opportunity (Horn, 2010a; Horn, 2013), impact on test scores (Horn, 2010a; Horn, 2013; Divényi, Sóvágó, 2021; Györfi, 2021), peer effects (Schiltz et al, 2019) and long-term overall effects on the society such as employment, wages and tertiary enrollment (Gurzó, Horn, 2015), on the other hand, the effect of test scores on later salary is also discussed (Hermann et al, 2022), but there is an important intermediate step that has been left out: the role of tertiary education. To the best of my knowledge, this research is the first to look at university outcomes.

Early selection in school systems tends to increase inequality because ability and socio-economic status are more correlated at early ages and the strength of this relationship diminishes later. The analysis of Horn (2010a) shows that elite programs tend to select applicants by ability, and since ability and status are strongly correlated at an early age, typically higher-status students are admitted to elite programs. In his study he exploits school admission rules in Hungary and uses age in months as an instrument: those students who were born between June and December 1995 have to start school in September 2002, and within this sample age in month and test scores are positively correlated, meaning that those students who were born earlier (exogen variation) perform better in school, which results in greater probability of admission to elite programs (outcome). The study also revealed that elite programs do have a positive impact on test scores: In grade 10, students attending elite programs perform better on the National Assessment of Basic Competencies (NABC) reading and mathematics tests than pupils in other types of school, even if we control for the 8th grade test scores and the effect of their social background. In other words, the test score gap between school types increases between the 8th and 10th grades. The two impacts together – that elite programs sort not merely on student competence but based on social backgrounds as well due to the early selection, and afterward students perform better in elite programs than their counterparts in other school types – result in increased inequality of opportunity. Similar results are documented in his later study (Horn, 2013). It seems that family status matters in the early selection since higher-status students are more likely to attend elite programs, ceteris paribus skills. He proposes a new instrument to measure the differences between value-added of school types - the distance between home and closest elite program – and finds that "early-selective tracks have higher value- added, but the size and the significance of this effect vary across cohorts and subjects" (Horn, 2013 p.38). The performance of students diverges in the 2 school types, and the study also shows that this selection is not a Pareto improvement, because those loose in society who are left in general primary schools when the best students leave the class, compared to those where there is no plausible option to leave class (closest school with elite program is too far).

Schiltz et al (2019) provide novel evidence on peer effects and show that high-achieving peers leaving the class (to go to these elite programs) have a mean negative impact on those who are left behind (those who did not go to elite programs). The effect is mostly driven by high-ability students left behind, by girls, and by high-SES students, although they do not find heterogeneous effects in sense of GPA, behavior, and higher education aspirations.

Divényi and Sóvágó (2021) apply a non-parametric bound methodology to investigate how enrollment in an elite program affects elite-school students' academic achievement in Hungary, measured with NABC test scores. They find a decrease in female and low-ability students' mathematics test scores two years after enrollment, which reverses later on, and becomes positive for each gender-ability group 4 years after enrollment. The positive effect is concentrated at the upper end of the test score distribution, and they are stronger for high-ability students suggesting that they benefit more from elite-school enrollment which can contribute to the persistence of inequality as well. In her ongoing research using RDD, Györfi (2021) also found heterogeneous returns measured by mathematics test scores: girls benefit less from competitive elite programs than boys do, although the results are inconclusive. Furthermore, in the RDD setting, estimates are only informative at the admission cut-off.

Horn (2013) assessed the short-term impact of elite programs on students' academic outcomes and found that they have a higher added value than regular schools, but a negative impact on the outcomes of students left behind (who remain in regular schools). Based on this result, Gurzó and Horn (2015) try to uncover whether the higher value-added also has a long-term impact by leading to higher post-secondary education and better labor market outcomes. The authors exploit that the establishment of elite programs in 1990s Hungary provides conditions similar to a natural experiment and found no overall increase in the average tertiary education attainment and completion of the Hungarian population, as well as no improvement in the average overall unemployment and wages. Zero results on the aggregate level may be caused by the fact that selection resulted in as much positive effect for students of elite programs as negative for students left in general schools. This study investigated a similar issue as Duflo, Dupas, and Kremer (2011), in a sense that both studies look at the overall effect of a kind of ability grouping in schools. However, Gurzó and Horn (2015) did not confirm the overall positive effect, contrary to Duflo, Dupas, and Kremer (2011). It would be an interesting topic of further study what might have caused the different results and what could be done so that the Hungarian system could also become advantageous for both groups.

I would also like to mention two studies, which does not directly relate to elite programs, but helps to understand the relationship between test scores, higher education attainment and wages. If elite programs are associated with higher test scores and educational attainment, than they can affect wages as well. Hermann et al (2022) investigated the effect of high school numeracy and literacy skills on longer-term outcomes: wages and employment around 25 years of age. Their results showed that the test scores of the NABC test in grade 10 are strongly correlated with early labor market outcomes. In line with international evidence, cognitive skills assessed in secondary school are good indicators of later labor market success. They have found that on average, numeracy (measured by mathematics test scores) has roughly twice as strong an effect as literacy (reading test scores) on early wages and employment. Cognitive skills influence the wages of people with higher educational attainment, although, for those with lower education, the test score does not affect subsequent earnings. The Admin3 database[6] that I use in this study is not long enough to investigate the impact of tracking on wages, but elite schools influence wages through higher education attainment (Estrada and Gignoux, 2017). The private average global return to a year of schooling is 9% a year (Psacharopoulos and Patrinos, 2018), and across OECD countries, the average internal rate of return to tertiary education is 15% for men and 19% for women, in Hungary, it is 17% and 12% respectively[7] (OECD, 2021). Accordingly, if elite programs influence higher education attendance, then it also induces an indirect impact on wages in the long run.

I plan to contribute to this literature by measuring the effect of elite programs not only on test scores and conventional post-secondary educational outcomes (university enrollment and completion) but also looking at heterogeneity in these outcomes (STEM track choice and quality of univeristies, which is of increasing importance in the time of higher education expansion), as well as heterogeneity in treatment effects (gender, SES, test scores, secondary school/teacher quality). Furthermore, I will also explore possible mechanisms: whether elite programs have an impact through changes in test scores, GPA, and college aspirations, or something else (for example teacher and school quality or peer effects).

2.3. Measuring universities

The main goal of this study is to discover how and to what extent elite programs influence university outcomes. In the previous chapter I described the elite programs in Hungary, which is the treatment variable in our model, on the other hand, different university measures will serve as outcome variables, therefore it is also important to understand how to measure higher education in Hungary – with special focus on its quality - which will be introduced in the following part.

The conventional outcomes to measure tertiary education are enrollment and completion. I also find it important to differentiate between institutes based on quality

[6] Admin3 is the linked administrative panel sample (random 50%) of the Hungarian population between 2008 and 2017 (Sebők, 2019). See more details in 3.1. Dataset.
[7] Values are based on the difference between people who attained tertiary education and those who attained upper secondary education.

and types of majors to get more detailed insight into post-secondary learning outcomes and success. I will examine the following outcomes:

- Enrollment: start attending a university.
- Completion: obtaining a degree.
- Field of studies (STEM, art, medical, law and governance).
- Quality of university: international rankings and bibliometric indicators.
- Enrollment in MA studies.

Enrollment at the BA level gives us a first insight into post-secondary learning outcomes, whereas completion of university studies (obtaining a degree) measures a somewhat different aspect since drop-out is also a relevant factor and university enrollment rates can differ substantially from completion rates. The latter determines the educational attainment level of individuals, which is an important measure of human capital accumulation and also has a signaling value in the labor market. These two are the most commonly used and most easily accessible measures of higher education outcomes. Although there are great differences between the field of studies and the quality of institutions, it would be interesting to see how elite programs influence the probability of choosing certain majors and prestigious universities. It could also contribute to a deeper and more detailed understanding of the effects. I will distinguish STEM majors since they could be of high policy relevance and look at differences in those fields, which are ranked high in students' preference rankings (art, medical studies, see university ranking in Csató and Tóth (2020) based on revealed preferences of applicants) and also those which lead to a traditionally prestigious degree and position of power (medical, law and governance). I also identify internationally recognized universities by synthesizing the performance of Hungarian universities in international rankings and looking at their publication performance based on Field Weighted Citation Impact (FWCI). These variables can show heterogeneity of the main outcome: enrollment. We can also observe a further layer of higher education, which is enrollment at the MA level, but the database is not long enough to observe any subsequent stages (for example PhD studies, or labor market transition).

The quality of the universities is of increasing importance in the time of higher education expansion. Rankings provide examples of how to measure quality, which helps to differentiate the outcome variable. I create a university quality indicator using various aggregate data such as international rankings, publications, citations, size of faculty, and preferences of applicants to obtain an outcome variable of greater variability. I obtained several bibliographic variables for the period 2008-2021 that can be used to measure university quality and often serve as a basis for international rankings (Falch et al 2022). I link it to the Admin3 database (see in 3.1. Dataset) to see not only whether an individual is enrolled in a university or not, but also the quality and value-added of that institution. I am going to examine the probability of further study at the MA level as well. The analysis will include several robustness checks related to the choice of variables and cohorts for the study.

In a recent study, Falch et al (2022) showed, that measure of scientific publications per full-time faculty at the institution – an indicator of faculty quality in terms of research orientation and research quality – is significantly correlated (r=0,507*) with the income-based value-added measure calculated by the authors for the higher education institutions in Norway. Another quality measure: full-time members of the scientific staff per student – an institutional input characteristic, traditional resource measure – also shows a significant positive correlation (r=0,365*), although to a lesser extent. These findings also justify the use of these indicators in the model.

2.3.1. International rankings, with a focus on bibliographic indicators[8]

In the following paragraphs, I will summarize the indicators used by main international and Hungarian rankings, with special emphasis on bibliographic measures. International university rankings place great emphasis on the measurement of scientific performance, while Hungarian university ranking practice is less focused on it, and unlike in the international literature, does not compute it based on publication performance, but rather based on the academic degree of staff. Using individual data based on the Hungarian Scientific Bibliography (MTMT), our study provided several possible examples of how an important aspect of scientific publication performance could be incorporated into Hungarian university rankings. We also showed that several different measurement techniques can be applied, but that they reach very similar

[8] This part builds on my previous research that measures scientific performance in Hungarian higher education institutions of economics and business (Csóka et al, 2019). The co-authors consented to the use of the research results in the dissertation. The chapter about the review of international and Hungarian rankings was almost entirely my own work, my overall contribution to the publication was 70%.

conclusions, meaning that a new criterion based on scientific publications gives a robust result. I would like to contribute to the integration of bibliographic indicators into Hungarian practice and raise awareness of the importance of research in higher education, in line with international trends.

The development of higher education rankings and the evaluation of institutions has been an area of intense research since the 1990s. Its relevance for public policy makers has been given by the higher education reforms. For example, from 1985 onwards, the British government proposed the creation of various performance indicators for universities to improve the effectiveness of higher education (Pollitt, 1990). And in 1995, the Chinese Ministry of Education launched a major initiative called Project211, to support the transformation of designated Chinese universities into world-class research universities to promote social and economic development. For higher education institutions, research on this topic was driven by the need to improve competitiveness, marketing objectives and optimize the allocation of funding. National rankings were also published during this period, starting with the US News ranking in 1987, the UnivPress ranking was launched in Hungary in 2001 (Fábri, 2016), and since the 2000s an increasing number of global rankings have been available, also showing the importance of the topic.

First, I present the international and Hungarian rankings, compare their focus and philosophy, and how it uses bibliographic indicators. I will then review the international ranking literature, with a particular focus on bibliographic indicators, followed by a review of the criticisms of ranking in higher education. Then, using the SciVal database (2022a), several possible metrics are shown to rank the institutes of Hungarian higher education. I examine the total scholarly output (number of publications), citations, value of awards, the average number of citations per publication, Field Weighted Citation Impact (FWCI), and quartile classification, as well as measures in proportion to the number of authors, if applicable. These indicators capture somewhat different aspects of academic performance, yet I find FWCI the best suitable to compare universities of different sizes and disciplines in Hungary. Correlation analysis shows that it is in a strong relationship with different indicators: it is highly correlated to citation per publication, output, citation, number of authors, and citation per author, furthermore, ranking based on FWCI overlap quartile rating as well. International rankings capture overall quality better than one single indicator (such as

FWCI) whereas they incorporate several different aspects of quality in higher education, so I decided to look at Hungarian universities in different international rankings: THE, QS, USNews, ARWU, and Scimago Q1-Q2 quartiles. I found a similar set of internationally recognized universities ranked in each case and constructed a dummy variable based on inclusion in these rankings which is considered a robust categorization of elite universities in Hungary.

Most of the global rankings are based on scientific publication and citation data but use a different methodology to measure research performance. ARWU's ranking only takes into account internationally recognized research performance: a university is ranked if it has any Nobel Laureates, Fields Medalists, Highly Cited Researchers, or papers published in Nature or Science, also, if it has a considerable amount of papers indexed by Science Citation Index-Expanded (SCIE) and Social Science Citation Index (SSCI) (ARWU, 2022). Bibliographic measures account for 70% of the index, they are based on the metrics provided by the Web of Science (WoS).

The US News, QS, and Times Higher Education (THE) rankings have a broader focus than the former ones, as they also assess the internationality of the university, its educational performance (QS and THE), and its reputation. A common feature is that the latter aspect is the most dominant in all three rankings: reputation surveys weight 25% for US News, 40% for QS, and 33% for THE. Among the research indicators, citations are the most relevant, and size-independent indicators are common. The total per capita indicator is calculated on a full-time equivalent (FTE) basis, with a specific methodology for THE and a one-third weighting for full-time employees and students and one-third for part-time employees for QS. US News uses the Web of Science database to calculate bibliographic measures, while THE and QS use Scopus. Of the global rankings examined, the teacher-student ratio is only included in THE and QS indicators, with a weighting of 4.5% and 20% respectively (THE, 2022; QS 2022; US News, 2022).

Scimago ranks not only universities but also other academic and research institutions. The index takes innovation, research, and web visibility into account. Research indicators account for 50% of the ranking criteria, and they are based on the publications registered in the Scopus database. For example, the number of publications registered in the Scopus database represents 8% of the final score, whereas high-quality Q1 and D1 publications have both a 2% and normalized citation impact is 13% (Scimago, 2022).

The Leiden ranking is also based on the WoS database, it assesses the impact, scientific collaborations, open access, and gender diversity of a university along dimensions such as the number and proportion of publications in the top 1%, 10%, or 50% by citation, the total and the average number of citations of publications, the number and proportion of articles co-authored with other countries or industry (Leiden Ranking, 2022). There is not one overall ranking, but users can sort institutions based on the different types of indicators individually.

U-Multirank's ranking philosophy is very different from the one presented so far, as it does not aim to create a single final league table with fixed weights, but is multidimensional, comprehensive, and user-friendly. Its flexible online interface allows the creation of personalized league tables based on more than 100 available indicators, including average citation rate, top cited papers, absolute and per capita (student and staff) number of publications (U-Multirank, 2022).

In addition, there are alternative ranking approaches, such as Webometrics, but these have less impact. Webometrics' ranking was originally designed to promote web presence, so it ranks mainly based on web content and links, based on the consideration that the web is key to the future mission of universities in the 21st century. At the current methodology, web visibility (number of external networks, so-called subnets linking to the institution's webpages) is 50%, whereas transparency (number of citations from Top 210 authors, excluding the top 20 outliers based on Google Scholar profiles) accounts for 10%, and the rest 40% is top cited papers in Scimago database (number of papers amongst the top 10% most cited in each one of all 27 disciplines of the full database). This is the largest ranking of higher education institutions, covering over 30,000 higher education institutions, while most other international rankings only cover between 1000-2000 due to strong inclusion criteria (Webometrics, 2022).

As far as Hungarian rankings are concerned, only UnivPress has comparable data, and Heti Válasz has regularly published higher education rankings, but its methodology has changed several times and its nature is rather business than academic. Both rankings focus mainly on indicators of incoming students, such as admission scores and language examinations. UnivPress also takes into account the qualifications of lecturers (habilitation, MTA title), as well as OTDK[9] competition places, PhD students, whereas Heti válasz looks at the gross income of graduates and the proportion of people employed in jobs requiring tertiary education (Fábri, 2016). The teacher-student ratio, or more precisely the number of full-time students per qualified teacher, is 13% in the UnivPress ranking, while this type of indicator was dropped from the Heti válasz after 2012. The adequacy of several of these indicators to measure university performance is questionable, but the most serious problem is that none of the Hungarian rankings use bibliometric indicators. This is a sharp contrast with international trends and with the guidelines for university quality assurance, accepted and applied both at home and abroad. In addition to these aspects, several possible approaches have appeared in the Hungarian literature, which mostly establish rankings based on the order of students' applications (Csató, 2013 and 2016; Csató and Tóth, 2020; Telcs et al., 2013).

I summarized the main focus of the international and Hungarian rankings, the type of bibliometric indicators they use and the sum of their weights altogether, and their data source in Table 2.

[9] Országos Tudományos Diákköri Konferencia (OTDK) in English: National Scientific Students' Associations Conference. It is the most prestigious academic competition in Hungary on university level, organized biannually.

Table 2: International and Hungarian rankings

Ranking	Focus	Bibliometric indicator	Database
ARWU	Only high-quality research (Nobel prize, citation, top journals)	top citations and publications (70%)	WoS
THE	Teaching, research, citation, international outlook, industry income	publication per faculty, citation (36%)	Scopus
QS	Reputation, faculty/student ratio, citation, international students, and faculty	citation per faculty (20%)	Scopus
US News	Research, reputation, international collaboration	average and top citations and publications (65%)	WoS
Scimago	Research, innovation, web visibility	top citations and publications (50%)	Scopus
Leiden	Scientific impact, collaboration, open-access publishing, gender diversity	highly cited publications	WoS
U-Multirank	Teaching and learning, research, international orientation, regional engagement, knowledge transfer	average and top citations, total and per capita publications	WoS
Webometrics	Visibility and impact on the web, top research	top-cited researchers and papers (50%)	Scopus, Google Scholar
UnivPress	Mostly attributes of incoming students, PhD students, OTDK, scientific degree of faculty, student/teacher ratio	-	-
Heti Válasz	Attributes of incoming students and their later labor market outcomes	-	-

Most international rankings include several publications measure, which is almost always compiled over a three- to five-year period, eliminating fluctuations and accounting for the length of the publication and knowledge creation processes, similar to the international literature (see, for example, Johnes, 1987; Chatzimichael et al, 2017). For global rankings, bibliographic indicators are usually generated using Scopus or the Web of Science database for international comparability, but for national rankings, the inclusion of publications in local languages other than English could also provide valuable information about scientific performance and quality (Csóka et al, 2019). In addition to the publications, citations are also a common indicator, as this can capture the impact of the research and the knowledge transfer that has been achieved. Publications in the top one percent of journals and with top citations also often appear, furthermore internationality, reputation, and web-based indicators are also included. Of the global rankings presented, only a few use the teacher-student ratio (THE, QS, U-Multirank).

University rankings have quite different approaches and methodologies, but publication and citation measures are essential in all of them.

2.3.2. Ranking literature and critique

In the beginning, research performance was expressed in terms of expenditure indicators because of data availability (Toutkoushian et al, 2003), and since the 2000s, bibliographic indicators have become more important: publications and citation-based indicators. The topic of research performance is an area of intense research, with earlier articles highlighting the importance of measuring publications, and more recent ones criticizing and refining the indicator - for example, by including citations alongside publications, or by weighting journals differently to include quality as a criterion alongside quantity (see Chatzimichael et al, 2017). They suggest focusing on quality indicators rather than quantity indicators and emphasize that these two concepts are not interchangeable (Pollitt, 1990; Johnes, 1987; Brooks, 2005; Tachibana, 2017; Bennett, 2001; Chatzimichael et al, 2017).

In terms of bibliographic indicators, both the rankings and the literature look at two different indicators. Sometimes the total number of publications is taken into account (size dependent), while in other cases the ratio per full-time equivalent (FTE) teacher is used to eliminate size bias. The literature tends to argue in favor of the latter (Bennett, 2001; Toutkoushian et al, 2003; Chatzimichael et al, 2017; Safón, 2013), but Brooks (2005) criticizes the per capita indicator for favoring efficiency over quality. When determining the denominator, i.e. the number of teachers, a problem may arise when counting lecturers who move between and affiliated with more universities and those who are teaching in different programs at the same time ("interdisciplinary activity" e.g. statistics, mathematics) (Johnes, 1987; Brooks, 2005; Moed, 2017), since in these cases a single lecturer contributes to the work of several different departments.

The international literature repeatedly shows that education and research can be complementary or even substitute activities, so caution is needed, as one can be changed to the detriment of the other (Johnes, 1987; Brooks, 2005; Toutkoushian et al, 2003; Chatzimichael et al, 2017; Johnes, 2018). In the case of the three missions of the university (teaching, research and development, service), it is not certain that an improvement in one of them will lead to higher quality in the other (Safón, 2013). A further difference is that there are input indicators such as the teacher-student ratio (Cakir et al, 2015; Bennett 2001) and output indicators such as the number of publications, furthermore, the teacher-student ratio is more of a feature of national rankings, while the number of publications is more of a feature of international rankings (Cakir et al, 2015, Johnes, 2018). Török (2008) also sharply criticizes the Hungarian rankings in this respect, as they do not take into account R&D performance indicators at all, which are important output and quality indicators, especially at PhD level (Toutkoushian et al, 2003). Török (2008) argues that meaningful rankings cannot be based on input indicators alone, since output performance is also largely determined by efficiency. Other indicators of educational quality often used in national and international rankings include student opinion, graduate retention, awards and recognition of faculty, and income factors (Pollitt, 1990).

International rankings have been criticized for being one-dimensional (Claassen, 2015; Moed, 2017) and containing several biases that tend to favor large, research-focused, English-language, science universities (Safón, 2013; Witte-Hudrlikova, 2013; Marginson, 2014; Fábri, 2016), thereby making value judgments between the different functions of the university (teaching, R&D, service), favoring research over teaching. Several authors highlight that there is no single universally accepted indicator or set of indicators that perfectly measures university (faculty) excellence (Bennett, 2001; Moed, 2017; Török, 2008; Johnes, 2018; Tachibana, 2017), as the mission of universities is very diverse (Witte-Hudrlikova, 2013). Nevertheless, the different rankings still have a strong impact on the behavior of institutions (Johnes, 1987; Johnes, 2018; Hazelkorn, 2007), so the design of an appropriate methodology is of utmost importance. Since there is no single, well-established methodology and weighting is arbitrary, some have questioned the use of composite indices (Marginson, 2014) and several papers have been written to correct the problem of weighting (Witte-Hudrlikova, 2013; Johnes, 2018).

2.3.3. Evaluation of Hungarian universities

In this chapter, I elaborate on what type of indicators could be used to measure university quality in Hungary, then I assess the Hungarian universities based on ranking indicators available in the SciVal database (2022a) and analyze the correlations between them. Thereafter I construct a dummy variable that shows which universities are included in the international rankings. Later, I will use this as an outcome variable in my model to measure university quality by distinguishing "elite" Hungarian universities that are internationally recognized. A similar distinction of high-quality universities is rare in the international literature relating to the effects of tracking – see for example Abdulkadiroglu, Angrist, and Pathak (2014) differentiate based on Barron's competitiveness indicator that shows selectivity of admission in the US, similarly to Ono (2001) in Japan, whereas Lu (2021) uses Russell Group universities that the equivalent of the American Ivy League of prestigious universities in the UK – so this can be a useful expansion of the existing knowledge about the effects of tracking on post-secondary outcomes.

First, I tried to construct a similar competitiveness indicator that captures the selectivity of institutions. In Hungary, students can apply for universities through a centralized application procedure where they can state their ordered preference of higher education courses, which includes the following pieces of information: name of the university, course (*szak*), whether it is full-time (*nappali képzés*) or part-time (*részmunkaidős képzés*), and financed privately or by public scholarship (*állami ösztöndíj*). Students got admitted based on scores achieved in a nation-wide centralized system, where their final exam test scores and GPA are accounted for. They can start at the bachelor's level (3-4 years of training) or indivisible master's level (5-6 years of training, for example, medical school is only available in this form). I used the statistics from Higher Education Admissions Database (FELVI, 2022) for the year 2014 (when most of the 6^{th} year cohort starts university), where the number of applicants and number of admitted students is available in the following categories:

- Sum of students.
- Full-time studies started at BA or indivisible MA training, in all forms of financing (both public scholarship and privately financed forms).
- Public scholarship studies at all levels.

In addition, applicants with first preference are also distinguished. I created several measures of admission rates (number of admitted students divided by the number of applicants) from these indicators, but none of them captured the quality of universities, instead the field of studies, capacities, and available public scholarship places drove the ranking. Universities from the field of arts (theatre, fine-arts, music) and theology performed the best since they are highly popular training compared to the capacities of those universities and the number of public scholarship places available. On the other hand, rural universities also performed better than would have been expected based on their student composition, due to higher capacities, which do not necessarily implicate higher student ability.

Another possibility to rank Hungarian universities would be based on applicants' scores, but as scores vary highly between fields of studies (for instance it tends to be higher for social sciences, and lower for engineering disciplines), normalization would be problematic when one aims to create one final league table. The difficulties are at least threefold: first, the same fields of study are available only at a handful of universities, so group averages are not too informative. Second, it is common that universities provide training in more fields (such as medicine, social science, law, and engineering within the same university) which makes the comparison harder. Third, the number of applications and applicants' scores are biased measures of quality, since they are highly dependent on the number of public scholarship places that influence applicants' behavior.

On the other hand, rankings incorporate several different aspects and capture overall quality better. Although it is hard to define what is a "good university", moreover the ranking literature is also abundant with methodological challenges (see chapter 2.3.2. Ranking literature and critique). International rankings apply different methodologies, but bibliographic indicators are essential in most of them, while Hungarian rankings are flawed since they do not include any bibliographic data and rather concentrate on input characteristics of students (see chapter 2.3.1. International rankings, with a focus on bibliographic indicators). Given the attributes of Hungarian higher education, it is a much smaller market than the mentioned US and UK examples, with around 50 universities providing BA or indivisible MA training, whereas there are roughly 5000 colleges and universities in the US and 100 in the UK, therefore quality should be measured differently, and I find “international recognition” a more plausible measure of quality in Hungary’s context.

In the following paragraph, I analyze the Hungarian universities based on internationally comparable data from SciVal (2022a). Table 3 shows Hungarian universities and the overview of their ranking indicators in SciVal (2022a)[10], sorted by Field-Weighted Citation Impact (FWCI). Information about quartile comes from Scimago ranking (which also uses Scopus bibliographic data such as SciVal): it is obtained by the institution in its country comparing the quartiles based on the overall indicator, research factor, innovation factor, and societal factor where not just universities, but also other research institutes are ranked (Scimago, 2022). Award value is in USD, and it makes sense to compare institutions with similar profiles (country, size, and discipline) based on this indicator. I distinguished size-dependent variables from size-independent (quartile, FWCI, citation per publication) metrics and prefer to use the latter ones to decrease the influence of size. Although the order based on the total number of output, citations, and authors seems similar to that of side-independent indicators, with only one exception in the composition of the top 5: Budapest University of Technology and Economics is second in total output, but performs lower in FWCI. The University of Pannonia and Eszterházy Károly University performs well in FWCI but has no award listed. In general, large universities with a medical focus perform best in the rankings. As publications from the areas of national importance (such as the law for instance) an from the field of arts and humanities are underrepresented in Scopus given its nature of international comparability, the performance of universities in these fields might be underestimated. Ranking order based on FWCI considerably overlaps with the quartile classification as well, which incorporates several different aspects and captures overall quality better.

[10] Metrics are averages in the timeframe 2011-2020, all types of publications are included and self-citations as well. There were no other options for this overview statistics.

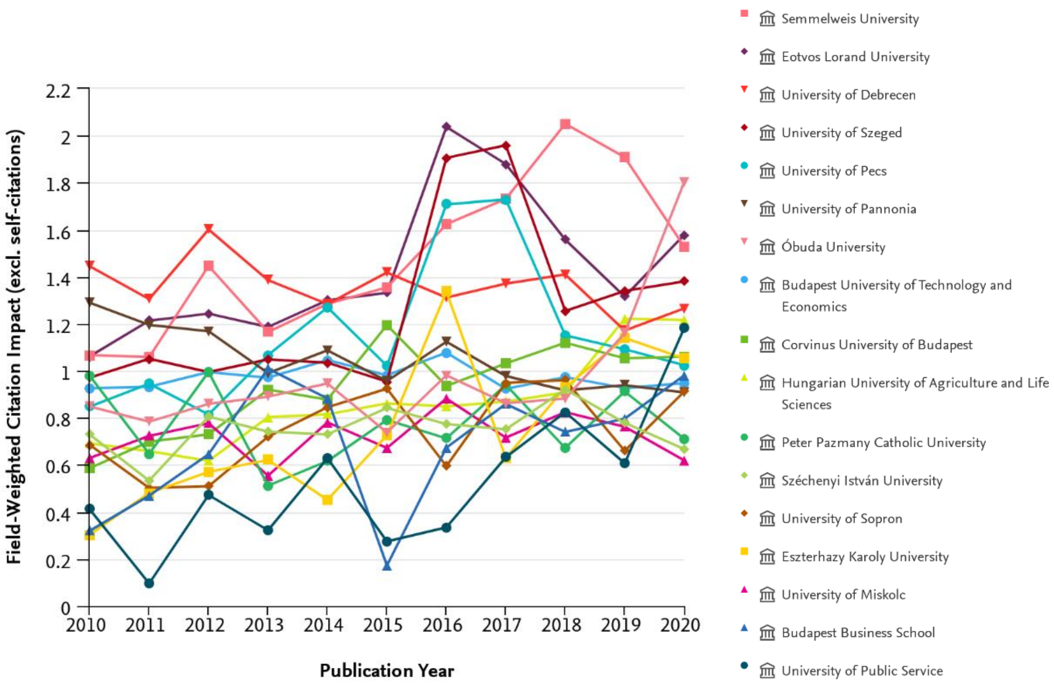
Table 3: Hungarian universities in the SciVal database

Institution	Quartile	FWCI	Citation/ Publication	Output	Citation	Award	Authors
Semmelweis University	Q1	1,56	24,7	12351	305319	16676887	6050
Eötvös Loránd University	Q1	1,49	21	15384	322689	80837318	5531
University of Debrecen	Q1	1,35	20,4	12247	249354	261251963	5535
University of Szeged	Q1	1,33	20	12416	247874	5304242	5305
University of Pecs	Q1	1,21	18	7521	135311	603856	3874
University of Pannonia	Q2	1,1	13,2	3425	45103	0	1273
Eszterházy Károly University	Q3	1,1	10	741	7428	0	264
Óbuda University	Q3	1,04	6,7	2748	18432	1323742	766
Budapest University of Technology and Economics	Q1	0,98	10,9	15166	165950	169984831	5262
Corvinus University of Budapest	Q2	0,97	11,1	2807	31151	114890	1380
Hungarian University of Agriculture and Life Sciences	Q2	0,88	11,6	4699	54466	26081277	2906
University of Public Service	Q3	0,82	4,7	364	1724	0	193
Budapest Business School		0,81	6,3	254	1599	0	130
University of Sopron		0,77	11,3	1194	13492	0	625
Péter Pázmány Catholic University	Q2	0,77	10,3	1463	15043	0	572
Széchenyi István University	Q3	0,75	5,8	1314	7610	583250	567
University of Miskolc	Q3	0,73	6,8	2455	16670	0	979

Source: SciVal, 2022a, Quartile comes from Scimago (2022).

FWCI is a useful measure to benchmark entities regardless of differences in their size, disciplinary profile, age, and publication-type composition, such as in our case: institutes within a country. Figure 6 shows FWCI in the period 2010-2020 for Hungarian universities, no self-citation is included, and I filtered for the “core” publication types: articles, reviews, conference papers, books, and book chapters. FWCI in SciVal indicates how the number of citations received by an entity’s publications compares with the average number of citations received by all other similar (same year, type, discipline) publications in the data universe: how do the citations received by this entity's publications compare with the world average? An FWCI of 1.00 indicates that the entity's publications have been cited exactly as would be expected based on the global average for similar publications; the FWCI of "World", or the entire Scopus database, is 1.00. FWCI of 1.9 indicates 90% above the world average (SciVal, 2022b). This metric is not reliable when an entity has a small number of publications, since outliers can skew the FWCI value substantially. It seems that the ranking is stable throughout the analysis period.

Figure 6: Field-Weighted Citation Impact and Year



Source: SciVal, 2022b.

To compare rankings based on different indicators, I conducted a correlation analysis on the Hungarian dataset. FWCI is highly correlated with other bibliographic measures, especially citation-related indicators, whereas awards (such as Awards/Author) and Output/Author seems to be different aspect. Variables proportional to author can be misleading because it only calculates with the authors who have Scopus-indexed research output, and not the total number of academic staff at the given university so it cannot be interpreted as average quality.

Table 4: Correlation of SciVal indicators

	FWCI	Citation/ Pub.	Output	Citation	Award	Authors	Output/ Author	Citation/ Author	Award/ Author
FWCI	1,00								
Citation/ Pub.	0,90	1,00							
Output	0,78	0,79	1,00						
Citation	0,89	0,91	0,94	1,00					
Award	0,36	0,37	0,65	0,54	1,00				
Authors	0,81	0,87	0,97	0,96	0,61	1,00			
Output/ Author	0,14	-0,12	0,13	0,02	0,11	-0,03	1,00		
Citation/ Author	0,93	0,93	0,84	0,91	0,42	0,84	0,22	1,00	
Award/ Author	0,34	0,35	0,64	0,52	1,00	0,61	0,10	0,40	1,00

Among these indicators, I find the quartile and FWCI the most suitable to show university quality in Hungary. FWCI is an appropriate indicator to compare universities of different sizes and disciplines, although there can be some outliers in the data, and it only considers scientific impact. On the other hand, the advantage of quartile is that this measure is constructed in a way that not only research, but innovation and societal factor are also taken into account, and it is appropriate for within-country comparison. Both indicators will serve as separate outcome variables in the empirical model to measure university quality.

I create a summarized ranking variable that is 1 if the given institute is listed in at least two of the following rankings: THE, QS, USNews, ARWU, in the period 2010-2020 (or closest data available). I chose this time interval because students in my sample could start higher education earliest in 2010, and most probably finish by 2020. Rankings often use 5-year averages, therefore the chosen period also seems reasonable. Based on these conditions, 8 universities were considered to be elite, and roughly 50% of students in the sample attended these universities. They are the following institutions:

- Eötvös Loránd University
- Budapest University of Technology and Economics
- University of Szeged
- Semmelweis University
- University of Debrecen

- University of Pécs
- Hungarian University of Agriculture and Life Sciences
- Corvinus University of Budapest

For the creation of this ranking dummy variable, I did not use the Scimago ranking, because 17 Hungarian Universities are listed there and roughly 90% of students in the sample attend these universities, so this kind of distinction would not highlight the elite universities well enough. Although I did observe, that the 2022 Scimago ranking also distinguishes quartiles Q1-Q4 for the first time in the history of this ranking, and if I have also used the list of only the Q1-Q2 universities for the construction of the ranking variable as well, then I would have got the same 8 universities as before. This might also indicate the robustness of the constructed ranking variable throughout different international rankings – not surprisingly, since they look at similar indicators in the broad sense, and they all use similar databases: either WoS or Scopus.

3. Data and Methodology[11]

3.1. Dataset

In this chapter, I present the data sources, method of sample selection for the analysis, and data transformations that resulted in the final dataset used for the calculations.

3.1.1. Data sources

The dataset consists of the linked administrative panel sample of the Hungarian population between 2008 and 2017 (Admin3). The sample selection is random 50% of the individuals who have valid social security numbers (TAJ) in Hungary, and datasets can be linked with an anonym identification code that is hashed social security numbers of the individuals. The education panel of the Admin3 is constructed by integrating student, class, and school-level information from the following datasets:

[11] The database construction and model specification outlined in this chapter (3. Data and Methodology) such as the descriptives and the baseline regression part presented later (in 4.1. Descriptive statistics and 4.4.1. Conventional measures: enrollment and degree) builds on the previously published preliminary findings of this research (Csóka and Horn, 2022). However, revised, updated and more detailed results are presented in this study. The co-author consented to the use of the research results in the dissertation. My overall contribution to the publication was 80%.

- National Assessment of Basic Competencies (NABC) measures the abilities of students in reading and mathematical literacy in grades 6, 8, and 10 in every Hungarian school. The database not only includes variables of school performance but background characteristics as well, which serve as control variables in the model. It also includes information about the school types: whether the student attended elite or general academic programs (treatment variable), and language preparatory programs (control).
- Registers for higher and for public education institutions, their students, and teachers come from the Educational Authority. The post-secondary outcome variables such as enrollment (BA and MA level), completion, and field of studies stem from the higher education statistics (FIR).
- The geographic location of students and schools from the Hungarian Academy of Sciences Geography (GEO) database makes it possible to construct distances that are used as an instrumental variable.

The data is confidential, although freely available for research purposes at the databank of the Centre for Economic and Regional Studies[12]. The Admin3 database provides a unique opportunity to analyze individual-level data on public education, higher education, competencies, and geographic location. The advantage of the database for my research is that it connects multiple data sources over a long period, using a consistent methodology. The existing database is complemented with further quality indicators aggregated at the university level.

Although most of the variables are raw test scores and survey data from NABC or administrative data from FIR, two special indicators were constructed for specific reasons: a family background index to incorporate SES into one variable, and one single primary school quality[13] measure that captures numerous elements of quality. The family background index is a continuous variable that has a mean of zero and a standard deviation of one, with a higher value indicating higher socio-economic status. The Educational Authority created this index on the individual student level from the number of computers and books in the household, the number of books owned by the

[12] See further documentation and information about the Admin3 database at the adatbank.krtk.mta.hu website and in Sebők (2019).

[13] I am grateful to János Köllő for providing his codes and data for the primary school quality indicator. The methodology and construction of the indicator is entirely based on his ongoing recearch with Luca Nagy.

student herself, and parental education and with optimal weights given by hierarchical regressions. The family background index (SES index) is a normalized variable within cohorts, therefore it has mean of zero and standard deviation of one, and condenses into a quantifiable, one-dimensional factor all the effects of the background variables that most determine student outcomes (Oktatási Hivatal, 2010 p. 54-56.). Furthermore, a specific school-level variable is constructed for primary school quality. Factor analysis identifies one single principal component factor that is the underlying latent factor of quality and relies on the following data on school-level averages in the period 2008-2012: test scores, the proportion of teachers without qualification, the proportion of students coming from a different region, student education aspirations, parental education, parental employment. The same calculation is conducted on data from 2013-2017. After factor scores are calculated, schools are categorized based on the quartile they belong to: high quality if they belonged to the upper quartile in both periods, middle if they were either q2 or q3 in both periods, low quality if they were in q4 based on the 2008-2012 and 2013-2017 factor scores as well and a school is labeled "uncertain" if there was a category change between periods. This results in a robust measure of quality throughout periods, which presumably holds for the period before 2008 as well (for which we don't have data).

In addition to individual-level data, I have obtained several higher education aggregate indicators that can be used to measure quality and often serve as a basis for international rankings, such as:

- Bibliographic measures from SciVal (2022a) such as field weighted citation index (FWCI), citation per publication, output, citations, awards, and number of authors available for 17 universities in Hungary (see Table 3: Hungarian universities in the SciVal database) complemented with quartile range from Scimago (2022)
- Ranking dummy variable, which distinguishes internationally recognized universities based on several international rankings (see the methodology of construction at the end of chapter 2.3.3. Evaluation of Hungarian universities)

I link it to the Admin3 database to see not only whether an individual is enrolled in a university or not, but also the quality and value-added of that institution. In Table 5 I list the variables, their meaning, and their type.

Table 5: Variables description

Dependent variables			
Variable	Meaning	Type	
Enrollment	Enrolled in university	Dummy (1: yes)	
Field of studies (e.g. STEM)	Enrolled in a major in certain fields of study (e.g. STEM)	Dummy (1: yes)	Quality
Ranking	Enrolled in an internationally recognized university[14]	Dummy (1: yes)	
FWCI	Field Weighted Citation Index of the university	Numerical	
Degree	Obtains a university degree	Dummy (1: yes)	
MA enrollment	Enrolled in university at MA level	Dummy (1: yes)	

[14] See definition at the end of 2.3.3. Evaluation of Hungarian universities.

Independent variables			
Language preparatory	Participated in language preparatory program	Dummy (1: yes)	Program type
Elite	Participated in an elite program	Dummy (1: yes)	
Mathematics test score	NABC score	Numerical	School performance
Reading test score	NABC score	Numerical	
GPA	Grade point average of last semester: 5 is best and 1 is worst	Numerical	
Aspiration	Plan to enter university	Dummy (1: yes)	
Female	Gender of the student	Dummy (1: yes)	Individual
Birthyear	Year of birth of the student	Numerical	
Primary school score	The underlying latent factor of quality.	Numerical	
Primary school quality	3 bins based on score: lower quartile, upper quartile, or middle.	Categorical	
SES index	Family background composite index	Numerical	Socioeconomic status
Cheap meal	Entitled to a cheap meal in school (state subsidy)	Dummy (1: yes)	
Free meal	Entitled to a free meal in school (state subsidy)	Dummy (1: yes)	
Free book	Entitled to free books in school (state subsidy)	Dummy (1: yes)	
Computer	Number of computers in the household	Numerical	
Car	Number of cars in the household	Numerical	
Bath	Number of bathrooms in the household	Numerical	
Book	Number of books in the household (binned values)	Categorical	
Siblings	Number of siblings living together	Numerical	
Parental education	Educational attainment level, separately for mother and father	Categorical	
Parental employment	Has a regular job, separately for mother and father	Dummy (1: yes)	
Parental age	Age, separately for mother and father	Numerical	
Help	Whether family helps with homework at least once a week	Categorical	
Distance elite	Distance from home to the closest school with an elite program by car (in kilometers between ZIP codes)	Numerical	Instrument
Distance university	Distance from home to the closest university by car (in kilometers between ZIP codes)	Numerical	

I define the treatment variable (elite program participation) with program type in grade 10, because dropout and repeat rates are very low in elite programs[15] and 10^{th} grade program type is the closest to grade 12 when students make their final decision about university attendance. Furthermore, program type is not available in grade 12, because there is no NABC test conducted at that grade. To control for differences in program types as well: I use a dummy variable on the individual level that shows whether the student attended a foreign language preparatory program, based on data that stems from the NABC survey. The proportion of students attending foreign language preparatory programs that provide 1 additional year of language education is different among general and elite programs (see at the end of chapter 2.1.2. Elite programs in the Hungarian education system), which can bias the results due to composition effect if participation in these foreign language preparatory programs affect university outcomes which seems a reasonable assumption.

3.1.2. The analytical samples

I have used two main samples for these results. The first is the sample of 6^{th} grade students in 2008 (Sample 6, baseline sample), the cohort that we can follow the longest from their enrollment into elite programs in 2008 until 2017, way into their higher education career. Another advantage of this sample is that pre-treatment control variables are available: test scores and other controls are measured at the end of grade 6^{th}, so right before they start the elite program[16]. However, very few of them finish higher education by 2017 so we cannot observe their completion (degree) and whether they continued their studies on MA level or not. I constructed another sample to be able to investigate these two important outcomes as well. That is the cohort in 10th grade in 2008 (Sample 10), where we can observe their university completion rates (degree) and MA enrollment with a much higher probability in 2017. On the other hand, the drawback of this sample is that their pre-treatment control variables in 6^{th} grade are not available, since the dataset starts only in 2008 when they finish 10^{th} grade. By that time, they are already 4 (or 6) years treated in elite programs. A panel sample is also

[15] Only 0,2% of students in 8-year long elite programs leaves or enters the program between 6th and 8th grade (Horn, 2013 p. 29).
[16] It is true only for the 6-year long elite programs, as the students of 8-year long programs have already been enrolled in elite program since 2 years at the time when their measurement happens at the end of 6th grade.

constructed to check for robustness of the findings about enrollment, where three cohorts of 6th grade students (who finish 6th grade in 2008, 2009 and in 2010) are selected since we can observe their enrollment until 2017, but this may not be true for later cohorts.

After filtering appropriate years and cohorts, I selected students who attend academic tracks: either 6- or 8-year long elite programs (whose secondary education begins earlier at grade 5 or 7) or 4-year long general programs (whose secondary education begins at grade 9). Students in vocational tracks are not considered in this analysis, since their attributes, interests, and motivations could differ substantially from students in academic programs, and these unmeasured factors could influence selection into treatment, hence would result in biased estimates. Individuals selected for this study all attend academic programs, so they are not a random set of students, but rather at the upper tail of ability- and socioeconomic distribution. It is important to bear in mind that the elite students are a selected group within the already selected students who attend academic programs. Sample sizes are around 18.000 in each cohort, with 20% of students in elite and 80% in general programs. Observations where all variables are nonmissing account for 77% of sample 6 (those students who are 6th grade in May 2008) and only 69% in sample 10 (those students who are 10th grade in May 2008), so I applied data transformations to increase the size of the samples used for analysis.

Table 6: Sample sizes for different sets of variables

	Sample 6	Sample 10
Total number of observations	18358	18626
Main control variables[17] nonmissing	14081	12813
After data transformation	17722	17514

Table 6 shows the importance of imputation, and how the sample size has changed after data transformation. In case of the baseline 6th grade cohort in 2008 the following numbers apply: among the 18358 students selected for the analysis who attend either elite or general programs, main control variables were nonmissing for 14081 observations, whereas after transformations that solved the problem of missing variables in most of the cases, the final analysis sample included 17722 observations.

[17] Gender, university aspirations, SES index, mathematics and reading test scores.

3.1.3. Data transformations

I conducted data transformations to alleviate the problem of missing variables and increase the size of the analysis sample that serves as the basis for the regressions. GPA and university aspirations were the two variables with the highest number of missing values: 4264 and 3084 respectively. Aspirations were already categorical, but GPA is numeric (accuracy to one decimal point), so it had to be transformed to categorical first. I created 4 groups with similar frequencies: low (below 4), lower-middle (4,1-4,4), upper-middle (4,5-4,8), and high (4,9-5). Five dummy variables were assigned to GPA, where the first 4 belonged to the 4 groups and the last took the number of 1 in case of missing GPA, and 0 otherwise (this last one is also called "missing flag"). As a consequence, I could use the whole sample in the estimation, not only observations with nonmissing GPA. Similar transformations were conducted for the set of control variables: first create categories from numerical values (although most of them were already categorical such as aspirations) then add missing flags. I omitted gender and test scores from the transformations and used their original values due to their low number of missing values and high importance.

There is a tradeoff between sample size and exact estimation of control variables coefficients. In those analyses where the latter matters more (descriptive statistics, analysis of sorting into elite programs and mechanism) I use the original variables to show the exact differences between factors and the magnitude of their effect sizes. Heterogeneity analysis makes sense just when the given variable is nonmissing, so it is also conducted on the smaller sample (a subsample where the relevant heterogeneity variable is nonmissing). On the other hand, in the regressions, the coefficient of elite treatment is of main importance which does not suffer from bias if I apply the transformed variables, but the internal and external validity of estimation benefits from the greater sample size.

3.2. The model

First, I will examine the characteristics of student composition in elite and general programs with descriptive statistics just to have a high-level overview of the raw differences, and I will analyze the mechanism of how sorting happens more thoroughly, where seems to be a consensus in the international literature as well as in research looking at the Hungarian context that early tracking age increases inequality (see literature review in 2.2. Impact of early tracking). On the other hand, results are mixed in terms of effects on academic achievement measured with short-term test scores and longer-term post-secondary and labor market outcomes. Therefore, I will look at the impact of elite programs on students' test scores first, then explore the effect on several types of university outcomes (enrollment, completion, STEM and other majors, quality of university, and MA enrollment).

Prior research suggests that socioeconomic status (SES) and gender influence a person’s educational ability, such as grades (GPA), non-cognitive skills, and values, hence educational outcomes (Brunello and Checchi, 2007, Scheeren, 2022a, b). I proxy ability by the NABC test scores, in addition, we can also observe (in the case of 6-year long elite programs) pre-treatment GPA and aspiration variables as well as socioeconomic status, gender, and primary school characteristics where students are coming from[18]. I argue that conditional on these individual variables selection into treatment is close to independent of the potential outcomes, therefore using the OLS or PSM will identify average treatment effects reasonably well. The identifying assumptions are unconfoundedness and overlap. I assume that we can observe all variables that have an impact on both enrollment to elite secondary school programs (T) and university outcomes (Y), meaning no omitted variables bias, therefore unconfoundedness holds. In addition, these observed variables predict well - but not perfectly - the enrollment in elite programs (overlap), so that average treatment effects can be identified using simple OLS and PSM methods.

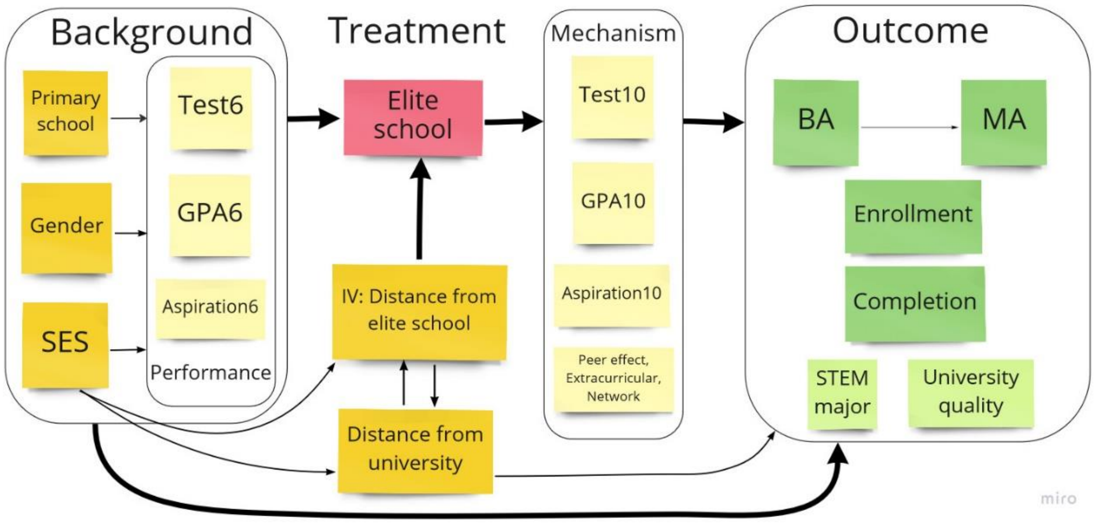
Unconfoundedness in social sciences is in most cases a strong assumption, and one cannot prove the lack of omitted variable bias. To overcome this obstacle, I also apply an instrumental variable.

The treatment is the participation in an elite secondary school program, which will be instrumented later on with the distance to the nearest school providing elite program. In the model, elite programs affect university outcomes through changes in test scores, grades, and university aspirations. I also plan to investigate peer, teacher, and network effects (enrolling into the same university as the ones before) in the future. The

[18] Due to the limited availability of primary school indicators – since these variables are unavailable for the students in 8-year long elite programs in baseline 6th grade cohort – I do not use it in the main regressions, but I show in chapter 4.7.1. Primary school quality, that inclusion of these variables does not alter the results.

measured university outcomes are enrollment and completion in BA and MA level courses, as well as the choice of STEM major, probability of changing courses, and university quality. See Figure 7 below for the causal map that clarifies the assumed mechanisms.

Figure 7: Causal map



I use OLS linear probability model with school fixed effects and PSM to estimate the effect of elite programs on higher education outcomes. I include various controls as outlined in Figure 7 above and estimate the following model:

$$Y_{ips} = \alpha + \beta T_{ps} + \delta X_{ips} + \gamma_s + \varepsilon_{is}$$

Where Y represents the outcome dummy variable (e.g. 10th grade test score, university enrollment and completion, STEM major) of student i in (elite) program p in school s, X includes controls on the individual level (eg. gender, socioeconomic status, and 6th grade school performance measured with NABC test scores, GPA, and educational aspiration), γ_s is school fixed effects measured in 10th grade and ε_i is the individual error term clustered at the school level. T stands for the elite program treatment: 1 if someone attended an elite program (either 8-year or 6-year long track), 0 if not. In the tracking period (5-12th grade) school type can only be determined in 10th grade for all types of programs, because in other years when NABC measurement takes place (6th and 8th grade) we still can't identify the control group: those students who attend general academic programs from 9th grade. But this should not mean a problem in this setting since the probability of track change (from elite program to general or the other way around) is close to zero.

The propensity score is a function of the relevant observed covariates (*X*): education aspirations, math and reading test scores at the end of the given grade, family background index[19] and gender. It is estimated in probit model and equivalent with the probability for an individual to participate in the treatment, given his observed covariates. PSM was conducted using the nearest neighbor method: for each treated observation one comparable observation was chosen in the control group, whose propensity score was the closest (nearest) to that of the treated one. After the nearest neighbor of each treated observation is identified, we can calculate the parameters of interest, by looking at the average difference between the matched treated and untreated individuals. The average treatment effect on the treated (ATT) is the most general estimate, whereas the average treatment effect (ATE) and the average treatment effect on the untreated (ATU) are calculated additionally. Matching also relies on the unconfoundedness assumption, such as OLS, and they both compare the average outcome for T = 1 (elite programs) versus T = 0 (general programs), conditional on observable confounders, but they do it differently. Matching detects and deals with the lack of common support since observations with propensity scores close to zero or one would not be matched by nearest neighbor matching, so "they estimate effects on the part of the data with common support only" (Békés and Kézdi, 2021 p.608). With common support, regression and nearest neighbor matching on the propensity score tend to give very similar estimates of the average effect.

In addition to these methods based on unconfoundedness, I address the assumption of no omitted variables by applying an IV estimation strategy (see section 5.1. Instrumental variables approach) as a robustness check. I use the distance from the closest school with an elite program as an instrument. This variable does not meet the exclusion assumption, since it can be correlated with the distance from the closest university and other socioeconomic factors, which can also influence the probability of university enrollment. Thus, I also condition on socioeconomic background and university distance in the IV models.

[19] Family background index consists of the following variables: parental education, books at home, and computers at home. See more detailed in the variable description of Oktatási Hivatal (2010, p. 54-56).

In our setting, it would not be possible[20] to link admission scores for elite programs (8-year or 6-year long track) to students, as student identification numbers (social security number) in the Hungarian secondary school admission database (KIFIR) are only stored for 2 years and deleted afterward due to data protection, so RDD – which is popular in the international literature - cannot be directly applied. Furthermore, RDD identifies the effect on the marginal student, and I am also interested in the average treatment effect that can be identified in OLS and PSM. Without a (quasi-) experimental design, causal identification is challenging, due to the strong assumption of unconfoundedness. From a methodological point of view, OLS estimates can suffer from selection bias due to unobserved variables that affect both the treatment and outcome at the same time (for example we usually cannot measure parental motivation). To overcome the possible shortcomings of the applied empirical strategy, the findings should be interpreted carefully, heterogeneity should be taken seriously and analyzed thoroughly, and it is worth conducting a reasonable amount of robustness tests to improve the external validity.

4. Results

In this chapter, I show the descriptive statistics and main OLS and PSM regression results. I follow the timeline of a student's education: first, I study the factors driving admission to elite programs and compare the role of test scores and social background, second, I analyze the elite programs' impact on test scores, third on longer term conventional post-secondary outcomes (enrollment and completion) and further indicators of university quality (STEM track, elite university, MA enrollment) as well.

4.1. Descriptive statistics

Table 7 summarizes the descriptive statistical results for the most relevant variables. The first two columns show sub-sample averages, and in the third column, there are differences and their statistical significance. The table includes the university

[20] Györfi (2021) tried to solve this problem and applied statistical matching on variables gender, date of birth and school ID to link the NABC test scores with the admissions data, although it is still work in progress.

enrollment rate, probability of STEM major, and elite university (based on international rankings) which are used later as outcome indicators, all of them are dummy variables. The explanatory variables come from the NABC survey data: gender, math and reading test scores, university aspiration (dummy variable indicating whether the student plans to proceed university studies), and several socio-economic indicators. The mathematics and reading scores are standardized to a mean of 1500 with a standard deviation of 200 in the 6th grade cohort, and each cohort and class is measured to this cohort, so one can compare the scores across years, within cohorts (Horn, 2013, p. 29). Parents' education is measured on an ordinal scale, and in the in the table the proportion of mothers and fathers with university degree is presented. Free meal[21] indicates the proportion of recipients of this social benefit. The book is a category variable, with 7 categories covering 0 to 1000 books or above, where 1 means less than a shelf of books (roughly 0-50 books), a value of 4 represents 150-300 books, and 7 means more than 1000 books in a given household. In the table the proportion of students with at least 150 books in their households is shown. We can also measure whether there is at least one car in the household, and the number of siblings living together.

University outcomes (enrollment, STEM major, and ranking) are higher for elite students, meaning that on average they enroll at a university with a 17,6 percentage point greater probability, and choose a STEM major with a 4,8 percentage point greater probability, in addition, they attend elite universities more often. The rate of female students and the proportion of those who get free meals is lower in elite programs. On average, those attending elite secondary school programs have higher test scores and academic aspirations, furthermore, also have a better family background. All differences between the two groups listed in the table are statistically significant at a 1% level (except for free meals, here the 10% level applies). The greatest differences can be observed in the parents’ education and SES index, which means that elite students have much socioeconomic status, and their parents have higher education.

[21] A type of state subsidy for children in Hungary, based on financial status of the family.

Table 7: Difference between elite and general secondary school programs (sample 6)

	Elite	General	Mean diff.
Math test (10th grade)	1828.332	1733.586	94.746***
Read test (10th grade)	1783.296	1713.542	69.754***
Enrollment	0.778	0.602	0.176***
STEM	0.326	0.278	0.048***
Ranking	0.664	0.608	0.056***
Female	0.541	0.591	-0.050***
Math test (6th grade)	1683.447	1594.523	88.923***
Read test (6th grade)	1670.596	1604.707	65.889***
Aspiration	0.934	0.819	0.115***
SES index	0.885	0.477	0.408***
Mother university	0.441	0.294	0.147***
Father university	0.386	0.240	0.145***
Free meal	0.017	0.022	-0.005*
Car in household	0.890	0.844	0.046***
Books in household	0.660	0.561	0.099***
Siblings	1.240	1.160	0.079***

Significant at $***$ 1%, $**$ 5%, and $*$ 10% levels. N=18358. Outcome variables are above the line: test scores as outcome are measured at the end of 10th grade. Control variables are under the line and measured at the end of 6th grade.

The descriptive statistics are quite similar for Sample 10 as well (see Table 8). Additionally, here we can also observe the completion rates (degree) and the MA enrollment rate. For these two additional outcome variables, we can find greater differences compared to the baseline probabilities, than in case of the first three. In this sense, the greatest difference can be found in MA enrollment, this probably indicates the rather academic style of the elite secondary school programs.

Table 8: Difference between elite and general secondary school programs (sample 10)

	Elite	General	Mean diff.
Enrollment	0.834	0.672	0.162***
STEM	0.349	0.305	0.045***
Ranking	0.647	0.598	0.049***
Degree	0.582	0.435	0.147***
MA enrollment	0.299	0.201	0.098***
Female	0.543	0.603	-0.060***
Math test (10th grade)	1842.766	1745.334	97.432***
Read test (10th grade)	1795.705	1729.479	66.226***
Aspiration	0.952	0.891	0.061***
SES index	0.907	0.488	0.419***
Mother university	0.409	0.292	0.117***
Father university	0.345	0.238	0.107***
Free meal	0.016	0.015	0.001
Cars in household	0.885	0.850	0.035***
Books in household	0.609	0.555	0.054***
Siblings	1.130	1.078	0.052***

Significant at $***$ 1%, $**$ 5%, and $*$ 10% levels. N=18626. Outcome variables are above the line, control variables are under and measured at the end of 10th grade.

Table 9 shows the mobility patterns from secondary education to university among students in the baseline sample (6th grade cohort).

Table 9: Transition from secondary school to university by program type (sample 6)

			University type	
	Number of students	No university	Non-elite university	Elite university
General	14591	39,8%	23,6%	36,6%
Elite	3767	22,2%	26,2%	51,6%

N=18358.

The proportion of students who do not enroll in university is higher among general students (nearly 40%) than elite students (22%), non-elite university enrollment is similar in the two groups, on the other hand, a substantial difference is present in the elite university participation rate, which is lower for general students (37% versus 52%). Percentage point differences between the groups (general and elite) are similar in 10th grade cohort, only university enrollment is roughly 7 percentage points higher. The p-value associated with the Chi-Square test statistic is close to zero, so we reject the null hypothesis that the two variables are independent. We have sufficient evidence to conclude that there is a statistically significant association between participation in an elite secondary school program and university enrollment, as well as enrollment into higher-ranking "elite" universities (university quality).

4.2. Selection into elite programs

In this paragraph, I study which factors drive admission to elite programs: what is the role of test scores and family background, and which is the stronger factor? The caveat of the database is that we do not have information about the applicants of elite programs, so we cannot filter who did participate in the admission process and who did not, we can only measure whether someone attended an elite program or not. The outcome is participation in elite programs, and I look at samples both with and without the 8-year long track. The analysis is conducted with standardized variables so that the effect sizes of SES and test scores can be compared to each other. Table 10 includes the coefficients of OLS (Model 1-6) and logit regressions. Female students have lower chances to get into elite programs holding either SES (Model 1) or reading scores (Model 3) constant, although gender difference disappears is we account for math test scores (Model 2, 4, 5, and 6). Furthermore, on an absolute scale, there are more female than male students in both cohorts: female students comprise 54% of elite students and 60% in general academic programs (see Table 7 and Table 8), but in comparison, their share is lower among elite than among general program participants. The coefficient of SES is always significant, although effect size decreases substantially if test scores are also taken into account. The effect size of mathematics test scores is similar to SES but greater than reading test scores, which suggests that admission to elite programs is influenced more by numeric skills than reading skills. In Model 5 distance of home from the closest school with elite program is accounted for, which explains away part of the SES effect, because the distance of home from elite program is correlated to SES (r=-0,27), which implies that more affluent families live closer to the schools providing elite programs. However, the inclusion of distance in the model does not change the coefficient of math test scores. One standard deviation higher math test score increases the probability of participation in an elite program by 5,3 percentage points. In Model 6 I filtered out students of 8-year long elite programs, which decreased the sample size but did not result in remarkable changes in effect sizes, although coefficients slightly decreased, and the reading test score became insignificant.

Table 10: Selection into elite programs: SES and test scores (sample 6)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Logit
Female	-0.017*	-0.011	-0.045***	-0.009	-0.008	-0.008	-0.053
	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.045)
SES	0.080***			0.057***	0.047***	0.034***	0.329***
	(0.003)			(0.003)	(0.004)	(0.003)	(0.025)
Math test		0.083***		0.053***	0.053***	0.040***	0.350***
		(0.003)		(0.004)	(0.004)	(0.004)	(0.030)
Read test			0.070***	0.016***	0.014**	0.006	0.108***
			(0.003)	(0.004)	(0.004)	(0.004)	(0.030)
Distance					-0.004***	-0.003***	-0.032***
					(0.000)	(0.000)	(0.002)
Constant	0.214***	0.212***	0.232***	0.207***	0.244***	0.175***	-1.210***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.040)
N	14470	17724	17730	14470	14267	12928	14267
R-Square	0.040	0.043	0.032	0.062	0.072	0.049	AUC: 0.695
AIC	14188	17408	17620	13864	13518	8761	13339

Significant at ∗∗∗ 1%, ∗∗ 5%, and ∗ 10% levels. Controls are standardized. Model 6: without 8-year long elite programs. Logit: log of odds ratios is reported.

Logistic regression shows similar results regarding the direction and significance of coefficients, furthermore, SES and math test scores have similar coefficients in this specification as well, which means the choice of functional form did not influence the findings. The odds of a student with one standard deviation higher SES is 1,4 times higher[22], ceteris paribus skills, and the same holds vice versa. The area under the curve (AUC) is an appropriate measure of performance in case of classification models (such as logit): it ranges between 0,5 and 1 and the higher value shows better accuracy of the model. If we take two random observations from separate programs (elite and general), there is a 69,5% chance that our model will correctly rank them. It is far from perfect prediction (the case when AUC = 1), but the explanatory variables in this model does explain part of the selection into elite programs.

Horn (2010) studied the selectivity of 8-year long tracks with probit and OLS models and found similar effect sizes: 0,027 coefficient of SES, 0,030 for math, and 0,020 for

[22] Odds ratio is calculated from the reported coefficients (log of odds ratios) in the following way: $e^{0,329}=1,39$.

reading test scores in the OLS model, although the sample included all 6^{th} grade students in 2008, not only the academic students as in this study and investigated the admission to 8-year long tracks not 6-year long. In a subsequent study, Horn (2013) investigates the admission to 6-year long elite programs and finds comparable, but higher odds ratios for SES and test scores in the logit model, and concludes that family status does matter, and one needs to have at least one favorable attribute (high status or high skill) to get accepted. The similarity of effects in these papers and my study implies that the selectivity of 8- and 6-year long elite programs do not differ substantially, and admission is influenced by both family background (SES) and school performance (test scores) at a similar magnitude, whereas within test scores differences in mathematics seem to be a stronger driver of elite program admission than reading.

4.3. Impact on test scores

In this section, I analyze the effect of elite programs on 10^{th} grade test scores, GPA, and university aspirations. I am looking for the answers to the following questions: do elite programs help increase students' test scores, GPA, and aspirations more than the alternative track (general programs)? Is this impact (if any) substantial or negligible? These are shorter-term education outcomes in 10^{th} grade so 6-4 years after the beginning of treatment, and in the later chapters I will also examine longer-term outcomes at the level of post-secondary education. Table 11 summarizes the coefficients of elite program participation for two types of models. In “raw” models’ the outcome is regressed only on the treatment variable, so these coefficients mean the raw difference between elite and general students’ outcomes in 10^{th} grade. In the second row, “conditional” models use school fixed effects and include the following transformed (see 3.1.3. Data transformations) control variables:

- SES variables: eligibility for state subsidies (cheap meals, free meals, and free books), living conditions and affluence (number of computers, cars, bathrooms, and books in the household), parental education, employment, and age (for both parents separately), number of siblings living together and whether family helps in homework at least once a week.

- Individual controls: year of birth[23] and dummy of language preparatory program type.
- School performance controls: GPA and university aspirations in 6th grade.

Further control variables are also included, without transformation:

- Gender, 6th grade math, and reading test scores.

In both models, standard errors in parentheses are clustered at the school level. The number of observations and R-squared in Table 11 belongs to the more preferred, conditional model. Looking at test score differences is important, it might explain why elite students continue their studies and get admitted to university.

Table 11: Elite programs and short-term outcomes (sample 6)

	Math	Reading	GPA	Aspiration
Raw	94,7***	69,8***	0,214***	0,101***
	(8,0)	(6,0)	(0,026)	(0,011)
Conditional, f.e.	21,1***	13,5***	0,055***	0,001
	(4,2)	(3,4)	(0,016)	(0,008)
Observations	16726	16731	13223	14229
R-Square	0,666	0,575	0,474	0,294

Significant at ∗∗∗ 1%, ∗∗ 5%, and ∗ 10% levels. Notes: f.e. is school fixed effects in 10th grade, and standard errors in parentheses clustered at the school level (521 schools). Controls in the OLS models: proxies for socioeconomic status such as cheap and free meals, free books, number of computers, cars, bathrooms, and books at home and variables of family background: number of siblings living together, how much help they get in preparation for school (learning, homework) at home, parental age, education, and employment. Also, individual controls: gender, year of birth, language preparatory class dummy, school performance: education aspirations, GPA, math and reading test scores at the end of the given grade. Imputation applied for all control variables except gender and test scores. Outcome variables are not imputed and measured at the end of 10th grade.

Significant raw differences are present in the case of all measured outcomes, which means that students' composition with regard to school performance is higher in elite programs on average. Much of this difference (roughly 75-80%) is explained away once pre-treatment characteristics are taken into account, which means that most of the difference in 10th grade is due to differences existing already in 6th grade. Furthermore, the effect on university aspirations even becomes insignificant in the conditional model. The effect on mathematics scores is greater than on reading scores, both effects

[23] Students born later within a cohort might have higher ability than their younger classmates at earlier ages, which might affect admission to elite program (T) and short-term education outcomes (Y) as well.

are significant, and comparable to 10% of standard deviation, which can be considered a sizeable impact. GPA effect size also significantly differs from zero and similarly around 10% of standard deviation. See summary statistics of 6th grade cohorts' 10th grade test scores, GPA, and aspiration in Table 15: Summary of results.

Presumably, students who try to get admitted to elite programs in grade 6 exert more effort on 6th grade NABC exams as well due to the preparation for the admission exam and this competitive period, on the other hand, students who actually get in (roughly 50% of the applicants) might exert less effort in 10th grade NABC test compared to their general secondary school counterparts, because for students in elite programs it is not their first NABC test in the class composition (they already had one in 8th grade), and the "ranking" of pupils within the class is already established, furthermore it is not a high-stakes exam so they can "wait" until the final exam (*érettségi*) in 12th grade to exert more effort. Given that the 6th grade test scores over-estimate, and the 10th grade test scores under-estimate the ability of students in elite programs, the difference in test scores between 6th and 10th grade is supposedly lower than the real increase in elite programs due to measurement error, under these circumstances the effect size might be downward biased. On the other hand, a bias in the opposite direction might be also present: students with higher motivations who are more eager to achieve success in education might self-select into elite programs, so the effect found is supposedly also upward biased.

4.4. Impact on university outcomes

I applied OLS and PSM models to investigate the effect of elite high school programs on several university outcomes. In the summary tables (Table 12: Empirical results: enrollment and , and further outcomes: Table 13, Table 14) I show how the effect size changes in the different model specifications – namely raw difference, fixed effects model, control variables model and the most preferred model including both control variables and fixed effects. I consider the specification using both control variables and fixed effects the most preferred one: on one hand, we need to condition on individual characteristics (control variables such as gender, SES, and before-treatment school performance) since these influence both the selection into treatment and the outcome variables, so omitting them would result in biased estimates due to endogeneity. On the other hand, school fixed effects control for various unobserved characteristics at the school level. Between-school effects (in the control variables-only model) contain not only the effect of elite programs, but also the differences between those schools that have both types of programs (elite and 4-year) and those that have only one type (4-year), so this result would be biased. I compare the results and highlight the most important findings.

4.4.1. Conventional measures: enrollment and degree

Table 12 presents the summary of the main regressions. The results in different samples (6^{th} grade in 2008 or 10^{th} grade in 2008) with different outcome variables (enrollment and completion) are shown in separate columns. The completion counts within 9 years after 10th grade (because the database is until 2017), so supposedly within 7 years after enrollment in the university[24]. The initial database covered 50% of the Hungarian population in the given cohorts, but since there are fundamental differences between school types in Hungary, in order to get a more precise estimate I have filtered out vocational schools from the sample, and analyses are run exclusively on academic schools: either usual academic 4-year programs or elite academic 6–8-year programs. The subsample of academically oriented students is where the identification assumption is the least troublesome, since these students are similar in the sense of academic orientation. That is, we are actually comparing students in general 4-year programs with students in 6-8-year elite programs. Results of 4 types of OLS model specifications are shown: either with or without control variables and school fixed effects. I used the exact same control variables as listed in the previous model (4.3. Impact on test scores). Fixed effects models control for all unobserved school-level characteristics, such as the composition of peers and teacher quality, although it would be a useful extension of this model to look at these factors also separately.

[24] In contrast to Western countries, in Hungary it is not a common phenomenon to take one year off after high school, students who are seeking to get a university degree mostly enroll in higher education right after the completion of 12th grade matriculation exams.

Table 12: Empirical results: enrollment and degree

	Y: Enrollment, Sample 6	Y: Enrollment, Sample 10	Y: Degree, Sample 10
Ymean [N]	63,8 [18358]	70,5 [18626]	46,5 [18626]
Ymean in general [N]	60,2 [14591]	67,2 [14873]	43,5 [14873]
Ymean in elite [N]	77,8 [3767]	83,4 [3753]	58,2 [3753]
β, OLS no controls, no f.e.	17,6*** (1,6)	16,2*** (1,4)	14,7*** (1,5)
β, OLS no controls, with f.e.	10,6*** (1,4)	9,1*** (1,2)	9,1*** (1,5)
β, OLS with controls, no f.e.	4,5*** (1)	3,6*** (0,9)	3,5*** (1)
β, OLS with controls, with f.e. [N]	3,5*** (1,2) [17722]	3,4*** (0,9) [17514]	3,4*** (1,2) [17514]
β, PSM with controls (ATT)	4*** (1,2)	3,4*** (1)	4,2** (1,6)
β, PSM with controls (ATU)[25]	3,4	4,6	1,7
β, PSM with controls (ATE)	3,5	4,4	2,2

Significant at $***$ 1%, $**$ 5%, and $*$ 10% levels. Baseline specification: f.e. is school fixed effects in 10th grade, and standard errors in parentheses clustered at the school level (521 schools in sample 6, and 530 schools in sample 10). Controls in the OLS models: proxies for socioeconomic status such as cheap and free meals, free books, number of computers, cars, bathrooms, books at home and variables of family background: number of siblings living together, how much help they get in preparation for school (learning, homework) at home, parental age, education, and employment. Also, individual controls: gender, year of birth, language preparatory class dummy, school performance: education aspirations, GPA, math and reading test scores at the end of the given grade. Imputation applied for all control variables except gender and test scores. Matching is on education aspirations, math and reading test scores at the end of the given grade, family background index (parental education, books at home, and computers at home), and gender. The number of treated/control groups in the 6th grade sample is (2868/2239), and in the 10th grade sample is (2471/1961)[26]. ATT is calculated using the nearest neighbor method with common support, where standard errors are bootstrapped (repetition=50).

Results are shown in percentage points.[27] Out of the 18358 students in the 6th grade sample, 14591 attend general and 3767 elite programs. Higher education enrollment is 60,2% in general and 77,8% in elite programs, hence raw enrollment difference is 17,6 percentage points. In other words, students who attend elite programs are on average 17,6 percentage points more likely to continue their education at universities than students in general programs (if those factors are excluded that influence enrollment in

[25] Command „psmatch2" used in Stata16 does not show significance for ATE and ATU estimates.
[26] In this subsample all matching variables (gender, math and reading test score, aspiration and SES index) should be nonmissing to be able to run the algorithm in a reasonable way, and each treated observation has one nearest neighbor in the control group, on the other hand, one observation in the control group can be the nearest neighbor of more than one treated observations, hence the smaller size of the control group.
[27] Coefficients are rounded to one decimal place for better readability, so they might not add up exactly due to rounding.

these elite programs, namely the control variables). When we look at differences within a given school (OLS fixed effects model), this difference decreases to 10,6 percentage points. This implies that there are significant within-school differences in the success of the general 4-year and elite programs. The most preferred OLS model uses controls and fixed effects as well: if someone went to an elite program instead of a 4-year long academic program within the same school, taking into account the observable selection by characteristics between programs within schools (holding other variables constant), than s/he has a 3,5 percentage point higher probability to enroll to university. In other words, an elite student with the same individual characteristics and attending the same school is 3,5 percentage points more likely to enroll in university than his or her general secondary school counterpart. The PSM algorithm yields similar results as the OLS controls model: average treatment effect on treated (ATT) estimate is 4 percentage points, furthermore average treatment effect on untreated (ATU) is 3,4 points and the average treatment effect (ATE) is 3,5 points. This implies, that the regression used has good properties: the assumption of the OLS linear function form does not have a large impact on the results, furthermore, the lack of common support in OLS estimation does not bias the results.

If we look at the differences between the model with controls and the model with controls and fixed effects, we can see that there are not great differences in terms of effect size and significance, for example, effect size decreases only 1 percentage point after controlling for school fixed effects in the enrollment equation for 6th grade cohort. This means, that the control variables capture the between-school variation in the outcome variable well, and omitted variables that relate to the sorting to schools do not substantially influence the results. Accounting for both pupil-level and school-level variables in the model results in over-controlling thus believed to provide a lower bound for the potential range of the real effect, although omitting school-level variables (unmeasured pre-existing differences between pupil groups) tend to overestimate the effects of more advantaged schools (Lu, 2021). This means, that the real effect size is supposed to be between the two last model specifications – which is fortunately not a wide range. Based on Card (1999), OLS estimates that control for family background may be downward biased, which provides another reason to presume that the estimate in the conditional model is a lower bound of the real effect size.

In the 6^{th} grade sample, there is no point to have a look at completion differences, because the students who were in 6^{th} grade in 2008 are expected to finish high school in 2014, whereas the database covers until 2017 and up until this year just 6% of the sample have obtained a university degree – in contrast to the expected rate of around 50% in that cohort.

Given that school performance variables in sample 10 are not pre-treatment, enrollment coefficients of 6^{th} and 10^{th} grade models are not directly comparable. To be able to examine completion rates, I have run the regressions on the 10^{th} grade cohort as well. In this cohort, we can compare the effect of elite programs on enrollment and completion. Elite programs increase the probability of enrolling in university by about 3,4 (both in OLS and PSM) percentage points in the 10th grade sample. The enrollment regressions in Sample 10 show only slightly lower effect sizes than the results in Sample 6 and even compared to the baseline sample mean (63,8% to 3,5% and 70,5% to 3,4%). As here I controlled for 10th-grade observable characteristics (test scores, GPA, aspiration, SES, and gender) I expected effect sizes to be much smaller, as elite programs increase test scores more than the general 4-year programs between 6^{th} and 10^{th} grade[28], and also have a positive effect on GPA (see Table 11: Elite programs and short-term outcomes). In other words, in sample 10 we are actually "over-controlling" for the effect of elite programs, so the fact that even in this sample they have a positive effect suggests that they do make a positive contribution to students' chances of continuing their education, not only by positive test score and GPA changes between 6^{th} and 10^{th} grades compared to general programs, but possibly through other channels and after 10^{th} grade as well. See a more detailed analysis of this mechanism in chapter “4.6. Mechanism”. The problem with the 10^{th} grade sample is that the control variables also stem from 10^{th} grade, 4 (or 6) years after the beginning of the treatment, meaning that they are already affected by the treatment.[29] Although we can argue that the elite dummy shows a positive association with the growth in most of the background variables that change over time during the treatment (test scores increase more and GPA decrease less, whereas aspirations decrease slightly more but its compensating

[28] The alternative to elite programs between 6th and 10th grade is primary school until 8th grade, and normal 4-year high school afterwards.

[29] 62% of the elite students attend 6-year elite programs, and the rest 38% attend 8-year elite programs and their 10th grade variables have already been under treatment for 4 years. In case of the 6th grade variables, 38% of the students are also already 2 years after the treatment, so that problem also persists in the 6th grade sample, but to a lesser extent.

effect is smaller than that of the test scores and GPA), so the resulting estimates are a lower bound of the real effect.

Nevertheless, the magnitude of the effect is nearly twice in the case of completion, in comparison with enrollment. Elite programs seem to increase university enrollment probability by around 3,4 percentage points (both in OLS and PSM). The effect of elite programs on university completion is similar in percentage points (around 3,4-4,2) but considering that the baseline probability is much lower for completing the university (46,5%) than to enroll in one (70,5%) the relative effect size of elite programs on completion is much larger than on enrollment. The reasons for this difference in effect size could be the subject of further research. Is it due to the higher value-added of elite programs that can be better utilized at the university education or some unobserved heterogeneity in the data? Elite programs might provide better teaching that prepares students better for the increased expectations at the university, in addition, they could be exposed to more positive peer influence and could already be better accustomed to higher workload, or their perseverance increased due to a more performance-oriented environment which could help them avoid dropping out at the university. On the other hand, the coefficient of completion could be upward biased if students who apply for elite programs are already more diligent, hard-working, and more determined to acquire a university degree before taking part in elite programs than their general program counterparts, and we cannot measure this pre-existing difference affecting university completion.

4.4.2. Further indicators: field of studies, quality of a university, MA enrollment

After analyzing the two conventional university outcomes (enrollment and completion) I look deeper and try to uncover some heterogeneity in the enrollment variable by distinguishing certain fields of studies (STEM, medical, art, law and governance) and the quality of the university in which the student enrolled, to gain a more detailed understanding of the effects. Quality will be measured with two indicators: whether the university is internationally recognized and based on its citation impact using FWCI which is an indicator normalized between fields and sizes of institutions. Lastly, enrollment in MA studies shows a higher level of university education and commitment to pursue further academic education and training.

In the samples that I used to examine the track choice of different fields of studies and the quality of universities just those students are included, who enrolled in university (N=11719 in sample 6, N=13102 in sample 10), although MA enrollment is analyzed on the whole sample 10 (N=18626) and only in this one cohort since MA enrollment is not available for sample 6. Furthermore, FWCI is only available for a subset of universities: those institutions that are listed in the SciVal database (17 institutions, see Table 3: Hungarian universities in the SciVal database).

Raw differences in Table 13 and Table 14 show, that significantly more elite students go to STEM, arts, and medical majors (although the latter is only significant in sample 6), and they go to internationally recognized universities with 5 percentage points greater probability, and also more likely to start MA level studies. On the other hand, there is no difference in the probability of law and governance majors and FWCI.

Once we control for observable characteristics and school fixed effects, differences between the probability of attendance of the examined majors disappear. In both samples, the raw difference in STEM track choice is around 5 percent, but after I control for the given set of background variables (SES, individual controls, and pre-treatment school performance) the coefficients became insignificant in OLS and matching as well. There is a similar pattern in medical studies. The mean of arts major in both samples is roughly one percent, and in sample 6 difference remains nearly the same even after including controls and school fixed effects, although this does not apply in sample 10. There is a significant difference in the probability of choosing law and governance majors between elite and general students in sample 6, but not in sample 10. In these two cases, it is possible that results in sample 10 underestimate the real effect since the controls used stem from grade 10 when students have already been "under treatment" (either in elite or in general programs) for either 4 or 6 years, and the coefficient only shows the effect in the last 2 years of studies in secondary school. Findings about arts and law and governance majors are inconclusive, but altogether it seems that even if elite secondary school programs have a significant and substantial impact on university enrollment and completion rates, they have no impact on which field of studies students pursue at universities. The lack of impact is likely because elite programs provide a comprehensive education with academic focus, and in most of the cases different specializations (for example advanced training in mathematics, natural sciences or literature) only appear at the last years.

Table 13: Empirical results: Field of studies, university quality (sample 6)

		Coefficient of Elite: β (s.e.)	
Outcome	Mean [N]	Raw difference	OLS with controls, with f. e. [N]
STEM	29 [11716]	4,8*** (1,1)	-1,4 (1,4) [11345]
Arts	1,2 [11716]	1,1*** (0,4)	0,9*** (0,3) [11345]
Medical	10,8 [11716]	1,9** (0,8)	0,8 (1,1) [11345]
Law and governance	6,7 [11716]	0,1 (0,6)	1,4* (0,8) [11345]
Ranking	62,2 [11719]	5,6*** (1,6)	2,9** (1,4) [11348]
FWCI	1,14 [10385]	0 (0,01)	0,01 (0,01) [10059]

Significant at ∗∗∗ 1%, ∗∗ 5%, and ∗ 10% levels. Baseline specification, see notes under Table 12.

Table 14: Empirical results: Field of studies, university quality, MA enrollment (sample 10)

		Coefficient of Elite: β (s.e.)	
Outcome	Mean [N]	Raw difference	OLS with controls, with f. e.
STEM	31,5 [13102]	4,5*** (1,1)	-0,1 (1,4) [12366]
Arts	1,1 [13102]	0,6** (0,3)	0,2 (0,4) [12366]
Medical	8,4 [13102]	1,1 (0,7)	0,3 (0,8) [12366]
Law and governance	7,5 [13102]	-0,6 (0,6)	1,1 (0,8) [12366]
Ranking	61,0 [13126]	4,9*** (1,6)	2,6* (1,5) [12389]
FWCI	1,13 [11582]	-0,01 (0,01)	0,01 (0,01) [10952]
MA enrollment	22,1 [18626]	9,8*** (1,1)	2,7** (1,3) [17514]

Significant at ∗∗∗ 1%, ∗∗ 5%, and ∗ 10% levels. Baseline specification, see notes under Table 12.

The next set of variables to differentiate in university enrollment is the quality of institutions (see Table 13 and Table 14). It is difficult to interpret the coefficient of FWCI. The mean outcome in both samples is around 1,1 which means that the number of citations received by publications of those Hungarian universities where the students from the sample enroll is 10% higher than the global average for similar publications since FWCI of 1 indicates the global average for similar (same year, type, discipline) publications in the SciVal database. It means basically that their scientific impact is slightly above the average. If we look at the effects, we cannot see any differences in either of the samples between elite and general students in terms of FWCI: raw differences and coefficients in preferred OLS models are both nulls. Although any indicator that tries to compare universities in one league table suffers from several methodological difficulties (see 2.3.2. Ranking literature and critique), normalization between fields of studies is problematic, and FWCI only looks at one aspect of university quality (citations), so we cannot draw any firm conclusions based on this one indicator. In my opinion, given the nature of Hungarian higher education where we compare far fewer universities than in the US and UK, international rankings provide a more reliable and robust measure of university quality than FWCI. Moreover, I find a categorical variable (internationally recognized or not) in this question more intuitive and better to interpret than a numerical variable (FWCI) which is the result of complex normalization, in addition, rankings also incorporate more different aspects of quality, not only citations. The proportion of students in the sample attending internationally recognized universities (shown by the ranking indicator) is roughly 60 percent in both cohorts, and 5 percentage points higher among elite students. This difference decreases slightly but remains significant 3 percentage points, which means that attendance in elite programs is associated with a higher probability of enrollment in a better university.

If we look at enrollment at the MA level, the difference between elite and general programs is substantial compared to the baseline probability: 20 percent of students in sample 10 continue their studies at the MA level, whereas it is 10 percentage points higher among elite students. Students who have similar observed characteristics in grade 10 within the same school, pursue further studies at the MA level with 2,7 percentage points higher probability if they attend elite programs.

To summarize the findings about the outcomes examined in this chapter, I did not uncover differences in enrollment in STEM majors, or highly preferred and prestigious medical studies, and found inconclusive evidence in preferred arts and prestigious law and governance studies. Similarly, no effect is identified on attending higher publication performance research universities (measured by the citation performance of the universities, FWCI) but elite students enroll in internationally recognized universities and MA level studies with a greater probability.

4.5. Compare short- and long-term educational effects

To sum up, the results, I have created Table 15 which includes the summary statistics of main outcome indicators: the number of observations, mean by program type (elite and general), the overall sample mean, standard deviation, the raw difference between elite and general academic secondary school students, finally the effect size in the most preferred OLS model that includes control variables and fixed effects as well. The first four rows (above the line) include results of short-term educational outcomes of elite programs, and the rest are longer-term post-secondary educational effects.

Table 15: Summary of results

	Sample	N	Elite	General	Mean	Sd	Difference	N	Effect
Math test	6	17291	1828	1734	1753	176	95***	16726	21***
Read test	6	17296	1783	1714	1728	153	70***	16731	14***
GPA	6	13667	4,13	3,92	3,96	0,65	0,21***	13223	0,05***
Aspiration	6	14702	0,92	0,82	0,84	0,37	0,10***	14229	0,001
Enrollment	6	18358	0,78	0,60	0,64	0,48	0,18***	17722	0,035***
STEM	6	11716	0,33	0,28	0,29	0,45	0,05***	11345	-0,014
Ranking	6	11719	0,66	0,61	0,62	0,49	0,06***	11348	0,029**
Degree	10	18626	0,58	0,44	0,47	0,50	0,147***	17514	0,034***
MA enrollment	10	18626	0,30	0,20	0,22	0,42	0,098***	17514	0,027**

Significant at ∗∗∗ 1%, ∗∗ 5%, and ∗ 10% levels. Models include f.e. and controls, see baseline specification under Table 12. The first N shows the number of students with nonmissing outcome variable, and the second N shows the number of observations in the models where the effect is calculated.

There are 18358 students in sample 6, and the number of observations in the baseline model for enrollment outcome is 17722 (see in the "Enrollment" row of Table 15). The number of observations can differ due to missing variables in the outcome variable (in case of test scores, GPA, and aspiration) or estimation on subsample – in case of STEM and ranking – because these analyses were conducted exclusively on those who enrolled in university. Estimates come from the 6th grade sample, except degree and MA enrollment where I used the 10th grade cohort (18626 students) because the database did not cover long enough time to look at these variables for the 6th grade cohort. Summary statistics of short-term outcomes show the 10th grade math, reading, GPA, and aspiration of the 6th grade cohort, so they differ a bit from both tables in 4.1. Descriptive statistics (there are 6th grade statistics of 6th grade cohort and 10th grade statistics of 10th grade cohort).

This way we can compare the significance and effect sizes of the different measures. If we look at the short-term outcomes, elite programs impact test scores and GPA positively, on the contrary, we rule out any effect on aspiration. The effect on math test scores is more than on reading test scores.

Findings are inconclusive about STEM track choice (and also about the probability of choosing other fields such as arts, medical, law and governance see Table 13 and Table

14), but elite programs have substantial effects on the quality of universities where students enroll, as well as on acquiring a degree and progressing to MA-level education. Compared to the respective baseline probabilities, the coefficient of elite programs is the highest in the case of MA enrollment, followed by degree, enrollment, and ranking, which means that the effect on BA completion (degree) and MA enrollment is greater in magnitude than on the simple BA enrollment rate. Short-and long-term educational effects are both positive, and effect sizes are comparable to 7-10% of standard deviation in the sample and always smaller than raw differences, which means that a substantial part of the variation is explained by observable pre-treatment characteristics (students with higher test scores and higher SES attend elite programs), but elite programs are associated with better short- and long-term educational outcomes even after controlling for these various individual and school-level characteristics.

The main takeaway of this section is, that the effects of elite programs on both short- and long-term educational outcomes are significantly positive (except university aspirations and choice of STEM major), and their size is comparable to roughly 7-10% of the standard deviation. It was also worth investigating not only BA enrollment and acquiring a degree, but also looking deeper and discovering the impact on the quality of the university to where students get admitted, and the probability of enrollment in MA level studies, to get a more detailed understanding of the effect of elite programs beyond the conventional higher educational outcomes.

In the next step, I look at the mechanism behind these findings.

4.6. Mechanism

I have studied the channels of the mechanisms, to find the reason of the success of elite programs: is it because they increase test scores and GPA of students, enhance motivation and aspirations, or something else that we cannot account for with the variables given in the model? To answer these questions, first I compare the statistics of these variables and their changes between 6^{th} and 10^{th} grade. Second, I change the value of 6^{th} grade mechanism variables (one at a time) to their corresponding 10^{th} grade value for all individuals and compare the regression results to see how the elite programs coefficient has changed due to the modification of the given variable, where the change in coefficient means the proportion of effect attributable to that channel.

Table 16 shows the means of university aspirations, GPA, and mathematics test scores and their changes between 6th to 10th grade. What we can see in the data, is that while test scores and GPA change more favorably (either improve more or decline less) among elite students than among their counterparts in general programs, aspirations are slightly declining among students and the decrease is greater in case of elite students compared to 4-year programs. The proportion of students planning to pursue studies at the university level decreased in elite (from 93,4% to 92,2%) and general programs (from 82,5% to 82,1%) as well, but more among elite students, by 1,2 and 0,4 percentage points respectively. The reason for this can be, that elite students study among higher-ability peers, hence due to this increase in comparison-groups (own classmates) quality they might start to value themselves less. GPA also decreased in both groups, but less for elite students. Mathematics test scores show greater growth for elite students. More favorable change (greater positive or smaller negative) is highlighted in green: students in general programs experience less decrease in aspirations, while the decline of GPA is more moderate for elite students and elite students both test scores increase more.

Table 16: Change in university aspirations, GPA, and test scores (sample 6)

	Elite			General		
Grade	6	10	change	6	10	change
Aspiration (Yes)	93,4	92,2	-1,2*	82,5	82,1	-0,4
Mean GPA	4,60	4,16	-0,44***	4,47	3,95	-0,52***
Mean math	1685	1829	143***	1597	1734	137***
Mean reading	1672	1784	112***	1606	1714	108***

Significant at *** 1%, ** 5%, and * 10% levels. More favorable change for elite students are highlighted in green, less favorable in orange.

I used the baseline model (6th grade sample and enrollment as outcome variable), because here we can see clearly the before-treatment (6th grade) and during-treatment (10th grade) values of the important school performance variables. In this baseline model, the coefficient of elite programs was 3,4%[30], and all control variables were measured before the treatment. The OLS coefficient presented is from the most

[30] To estimate this baseline coefficient, I used the original values of gender and school performance variables without data transfromation (see 3.1.3. Data transformations), which means the only differnece compared to the samples before is that GPA and aspiration should be nonmissing, so model estimated on N=13742 observations, and it gave 3,4% effect size. There is 0,1 percentage point difference compared to the 3,5% that is shown in the most preferred model where N=17722, only difference is the number of observations due to nonmissing GPA and aspirations, which seem not to alter the coefficient substantially, but still results in slightly different estimate.

preferred model, which includes controls and school fixed effects as well. I did the following: I always changed one certain variable or a given set of variables to its 10^{th} grade value instead of the original 6^{th} grade value. The change of the elite program's coefficient in this case means the size of the effect that can be attributed to that given variable that has changed. For example, in the model where I used all variables the same as in the baseline model, except I used the mathematics test score in 10^{th} grade instead of 6^{th} grade, the coefficient of elite programs became 2,6%. It means a 0,8% decrease in the effect, so around 24% (0,8/3,4) of the elite programs' effect on enrollment is resulted by the growth in mathematics test scores between 6^{th} and 10^{th} grade. The part of the effect explained by the reading test scores is 0,4% which is roughly half the size of mathematics test scores, and the two tests together account for nearly 27% (0,9/3,4). The GPAs of students in elite programs have changed in a more favorable direction (decreased less) compared to the GPAs of students in general 4-year programs on average, and it resulted in an even greater effect size than the mathematics test scores: 1,2% decrease in coefficient, meaning that 35% (1,2/3,4) of the elite programs' effect can be attributed to changes in GPA, what is more, coefficient became insignificant so it explained away the effect. Aspirations have changed in the opposite direction, less favorable for elite students than their counterparts in general programs, meaning that the aspiration effect of elite programs hinders elite students' chances to get into university, although not to a great extent (0,2 percentage point effect). Hence, if I use the 10^{th} grade values for all mechanism variables (2 test scores, GPA, and aspiration) then the coefficient decreases altogether only with 1,1% because of the factors showing in the opposite directions. In this model, the original 3.4% effect is reduced to 2,3 percentage points, which is comparable to the coefficient in sample 10 (3,4 percentage points) where only 10^{th} grade school-performance variables are observed instead of the pre-treatment 6^{th} grade ones, which implies that the enrollment coefficient in sample 10 is the lower bound of the real effect which we could identify if we would have available pre-treatment (measured in 6^{th} grade) variables for those students.

Table 17: Mechanism variables and their effects (sample 6, enrollment)

Model	Baseline	Math	Reading	Tests	GPA	Aspiration	All	Tests, GPA
OLS β	3,4**	2,6**	3**	2,5*	2,2	3,6***	2,3*	1,6
Effect	-	-0,8	-0,4	-0,9	-1,2	0,2	-1,1	-1,8

Significant at $***$ 1%, $**$ 5%, and $*$ 10% levels. Models include f.e. and controls, see baseline specification under Table 12. Additional information: each of the mechanism variables examined in separate models is measured in 10th grade, instead of 6th grade. Imputation was applied for all control variables except gender and mechanism variables (test scores, GPA and aspirations), Baseline N=13742.

The Main channels are math test scores and GPA. If I keep the 6th grade aspirations and use the tenth-grade scores for only the other variables (math and reading scores and GPA), the original 3.4% effect is reduced to 1.6%. Thus, roughly half of the impact of the elite programs comes from improvements in test scores and GPA between 6th and 10th grade. In other words, the 1,8 percentage points decrease in the coefficient means that we could account for more than half of the original 3,4% effect. The rest of the effect can be due to several reasons. For example, something happening after grade 10 that we cannot measure (such as a further increase in elite students' test scores and GPA compared to their counterparts), or can also be due to measurement error, whereas NABC tests are low stakes tests and there could also be different incentives for elite and non-elite students to complete this test (although it is highly unlikely within the same school). Other sources of the impact of elite programs can be peer-, teacher-, and network effects (enrolling into the same university as the ones before), or parental resources that we cannot account for in this model.

Elite programs in Hungary incorporate a form of earlier tracking and within-school ability grouping as well. Therefore, its impact can stem from two types of differences:

1. Difference between lower-secondary and upper-secondary level (in grades 7-8 or 5-8): students admitted to elite programs attend upper-secondary school for a longer period (6-8 years instead of 4), and upper-secondary programs are better preparation for university than lower-secondary primary schools which is the alternative in grades 7-8 or 5-8.
2. The difference within the upper-secondary level (in grades 9-12): elite programs provide better quality education indeed than their general counterparts in grades 9-12.

It is hard to distinguish which one of the above causes the identified effect, but regressions using 10th grade variables provide suggestive evidence that quality effects should be present. Even if I use the 10th grade school performance variables in the regressions, the impact of elite programs on enrollment stays positive and significant which means that there are differences between general and elite programs in the last two years of secondary school.

The reason for more favorable changes in GPA, mathematics, and reading test scores in elite programs (compared to 4-year programs) is the subject of further research: that could have been resulted by probably better-quality education (assigning better teachers within a school for this high-achieving elite classes), different curriculum, or peer effects. Several examples of these types of effects are found in the literature, for example, Brunello and Checchi (2007, p. 795) list the following ones: peer effects might be important, such as teachers' quality as we can also "suppose that teachers prefer to teach relatively high ability classes, […] assume further that better teachers have priority in the allocation to classes, […] then teacher quality is higher in classes with higher average student ability." On the other hand, there could be differences in school resources between schools and in curricula across tracks.

Pop-Eleches and Urquiola (2013) have shown that in Romania better teachers with higher certification standards are more likely to work in better-ranked schools, and this effect also persists within a school as one moves from a weaker to a stronger track, and even within tracks as one moves from a weaker to a stronger class. Varga (2009) also found evidence of similar behavioral responses of teachers on student stratification, as better teachers (higher educational attainment and experience) sort into schools with better students (more advantaged background) in Hungary, with the difference that her study analyzed primary schools (until 8th grade) and between school sorting, and not within school sorting at secondary school level as in our setting, although there is overlap with our setting in grades 4-8. One can suppose that a similar mechanism could be found there as well, as the composition of students also varies and is linked to teacher quality. Hermann and Horváth (2022) calculated the value added of teachers measured in test scores gain of their students in nearly 100 Hungarian secondary schools and discovered a within-school sorting of higher value-added teachers into classes with higher SES students. Clark (2022) studied the 1947 British reform and discovered that around that time "fewer than 10% of lower-track teachers had a university degree compared to almost 80% of higher-track teachers" (Clark, 2022 p.12) and highlighted those differences in "teacher characteristics likely mirror other resource differences (e.g., in buildings, books and equipment)". He also found differences in curriculum, suggesting that lower-track schools were discouraged from traditional academic courses and had more practical components, although government guidance prohibited narrowly vocational courses.

On the other hand, peer effects in educational settings are an area of intense research with large literature. Examples include Hoxby (2000), Hanushek, Kain, Markman, and Rivkin (2003), Angrist and Lang (2004), Hoxby and Weingarth (2006), Lavy, Silva, and Weinhardt (2012), Ammermueller and Pischke (2009), Duflo, Dupas, and Kremer (2011) and Schiltz et al (2019). Most of the studies find positive effects of high-achieving peers and negative effects of bad peers, and some find moderate or no effect at all. Although in his literature summary Sacerdote (2014) emphasizes that peer effects are highly context-specific and concludes that there is not yet a reliable consensus reached in this matter, however, he finds a pattern across studies which suggests that long-run social outcomes (e.g., crime, drinking behavior) and career choices show larger peer influences than do test scores, and effects are most probably non-linear and heterogeneous. Abdulkadiroglu, Angrist, and Pathak (2014, p.141) highlight that "the likelihood of omitted variables bias in naive estimates motivates much of the econometric agenda in this context". The paper of Schiltz et al (2019) has special relevance for my study given the Hungarian context: they find high achieving peers leaving the class (departing for elite programs in Hungary) has a negative impact on those who are left behind in primary schools, the mean negative effect is especially driven by high-ability, high SES students and disadvantageous for girls.

4.7. Heterogeneity

I analyze heterogeneity to see whether there are differences in the magnitude of effect between groups based on gender, socioeconomic status, education aspirations, test scores as well as school-level test-score value added (school quality) to get a more detailed description of what drives the results. I estimated the same model but separated the sample into "advantageous" and "disadvantageous" groups based on the heterogeneity variables one by one, for example first I estimated for the subsample of male than female students, second, I took the subsample of students who initially (in grade 6) planned to go to university than the ones who didn't. Table 18 shows the coefficient of the elite program in the different subsamples. Students who have at least one parent with a university degree belong to the group "parent with degree", students who meet the following criteria: upper quartile in math and reading test scores and above median GPA belong to "good student", whereas "good math" is defined as the upper quartile of ability distribution based on mathematics test scores. The two subsamples are always complementary sets of the whole sample, although sometimes they do not add up to 17722 sample size, since heterogeneity analysis is not conducted on observations with missing values in the given variable (for example, there are 7218 high-SES and 7252 low-SES students in the sample, and for 3252 students SES index is not available).

Table 18: Heterogeneous effects in subsamples (sample 6, enrollment)

	Coefficient of Elite: β (s.e.)[N]			
Subsample	Advantaged	Disadvantaged	Subsample	Diff
Male	2,4 (1,8) [7402]	4,9*** (1,5) [10320]	Female	-
Plan university	2,2* (1,3) [12843]	10,5** (5) [2408]	Not plan university	**
Parent with degree	3,1** (1,3) [7264]	4,9*** (1,7) [10458]	Parent without degree	-
Good student	2 (2,2) [1582]	4,0*** (1,5) [12489]	Not good student	-
Good math	3,1* (1,6) [4432]	4,2*** (1,5) [13290]	Not good math	-
SES > median	2,2* (1,3) [7218]	6,6*** (2,4) [7252]	SES <= median	**

Significant at *** 1%, ** 5%, and * 10% levels. Models include f.e. and controls, see baseline specification under Table 12. Additional information: each of the heterogeneity variables examined in separate models should be nonmissing in the respective model. The significance of the difference between the coefficients for advantaged and disadvantaged subgroups is reported in the last column.

There are positive effects in all cases, although with a higher p-value (lower significance) for advantaged groups. The effect is even insignificant for males and good students, meaning that they would enroll in university with a similar probability even if they would not attend elite programs. Although it does not necessarily mean that elite programs are not beneficial for them, but insignificant estimate might be also due to decreased size of the given subsample. The analysis conducted on separate subsamples shows that students with low initial university aspirations benefit from elite programs the most. Low-SES students also experience a high increase in university enrollment probability (6,6 percentage points), furthermore, attendance of elite programs for pupils whose neither parent has a tertiary education and for female students result in a 4,9 percentage points higher probability of university enrollment, ceteris paribus. To test the equality of regression coefficients, I used the formula proposed by Paternoster et al. (1998). The difference between the coefficients of the subgroups is significant in cases of university aspirations and SES. For other heterogeneity variables I cannot reject the null hypotheses of equal coefficients of advantaged and disadvantaged subgroups. It means, that those students who initially (at the end of 6th grade) did not plan to attend university benefit significantly more from elite programs than those who did. Similarly, students with SES under the median experience higher gains than those, who have above-median SES. For gender, parental background, and initial test scores the evidence about the differences between the subgroups is inconclusive.

There seems to be an equalizing effect between students attending elite programs since findings imply a greater positive effect for more disadvantaged students: low-SES, lower school performance, lower aspirations students, and girls. Although sorting into elite programs harms equality in the society since it is not merely based on ability but on SES as well (which is linked to test scores if sorting happens at earlier ages), equality of outcomes is increased among students admitted to elite programs.

4.7.1. Primary school quality

Primary school quality is an important factor in this study, although I did not use it as a control variable in the main models due to data availability: my main sample consists of students who were 6th grade in 2008, and among them, we do not have information about which primary schools those students attended who participate in 8-year long elite programs, since they left primary school in 2006 when no data is available. Given that constraint, I need to restrict the sample to students of 6-year long elite programs and general 4-year long academic programs and filter out students of 8-year long elite programs if I use primary school quality. Therefore, this paragraph also functions as a "transition" to the next chapter (5.2. Robustness) because I will not only conduct heterogeneity analysis and compare the effects of elite programs in subsamples differentiated based on primary school quality, but also compare the baseline model with the one on a restricted sample (without 8-year long elite programs), with and without primary school quality as a control variable.

Students of elite programs come from better primary schools. Table 19 shows the distribution of primary school quality among general and elite students: the proportion of students coming from high-quality primary schools is higher among elite students

(63%) than among students in general programs (53%), whereas only 7% of elite students come from low-quality primary schools which is lower compared to 10% in general programs. The chi-squared test implies that primary school quality is not independent of elite program participation (Chi-squared = 70,92 and p-value=0,000), elite secondary school students come from better primary schools, on average.

Table 19: Distribution of primary school quality among general and elite students (sample 6)

Primary school quality	General	Elite	Average
High	53%	63%	55%
Middle	37%	30%	36%
Low	10%	7%	9%
Total	10965	1894	12859

Additional information: sample without 8-year long elite programs.

1764 students were filtered out from the sample (8-year long elite program students) because of no available information about which primary school they attended, so the sample consists of 16594 students of 6th grade cohort in 2008. Primary school is labeled "uncertain" for 1631 students and missing for 2104 observations in this sample. The elite coefficient in the baseline model (effect on university enrollment) was 3,5 percentage points, and in the sample without 8-year long elite programs, it became 3,1 percentage points. If we control for pre-treatment primary school quality with the category variable constructed based on factor score quartiles, the elite coefficient does not change, which means that even though the sample is not balanced on primary school quality and elite students attended better primary schools, the difference in this pre-treatment characteristic is probably already controlled for due to the presence of other control variables, such as for example test scores and SES of students. If we use the factor score instead of the category variable, the effect becomes 3 percentage points which is not a substantial change either. If we include primary school quality variables in the model, we might undercontrol for the primary school effect and not grasp every aspect of it. In order to make sure that the significant positive effect of elite treatment is not due to selection from good primary schools, I included a second type of fixed effect in the baseline model: not only secondary school (in 10th grade) but primary school (in 6th grade) as well. This way we overcontrol for the primary school effect, but the coefficient remains significant and similar magnitude (2,8 percentage points). Given that the elite coefficient is similar is all specifications, it seems that the elite coefficient is not driven by sorting of students from better primary schools. Table 20 summarizes the regression results, and also includes estimates on subsamples.

Table 20: Primary school quality and heterogeneity (sample 6, enrollment)

Model	Coefficient of Elite: β (s.e.) [N]		
Baseline, without 8-year elite	3,1**	(1,5)	[16027]
Conditional on: category	3,1**	(1,5)	[16023]
Conditional on: factor score	3,0*	(1,5)	[13972]
Baseline, with primary school f.e.	2,8*	(1,6)	[15758]
Subsample			
High quality	2,3	(2)	[6737]
Middle	4,2	(2,9)	[4486]
Low quality	7,2	(7,1)	[1190]

Significant at *** 1%, ** 5%, and * 10% levels. Models include f.e. and controls, see baseline specification under Table 12. Additional information: sample without 8-year long elite programs.

Results show that the effect is heterogeneous based on primary school quality: attendance of an elite program increases the probability of enrolling in university more for students coming from lower quality primary schools (7,2 percentage points) than for students from better primary schools (2,3 percentage points). However, these coefficients are insignificant, which is probably because of the decreased sample sizes. Students of better primary schools benefit less from elite programs than those who attend less prestigious primary schools. This implies a similar equalizing tendency as the previous heterogeneity analyses between advantaged and disadvantaged groups (see in Table 18).

5. IV Estimation, Robustness

In this chapter, first, I address causality and test whether the results also hold in IV estimation. If not, that could be the sign of a possible endogeneity problem. In this case, endogeneity can be caused by either a possible omitted variable or measurement error, but other sources of endogeneity (simultaneity) are highly unlikely. The NABC database is rich in background variables, however, a possible omitted variable could be for example competitive mindset of the student or parental motivation, although even if it is not explicitly measured it is presumably proxied by parental education and aspiration of students. Measurement error can occur if the test scores do not show the real ability well enough. Possible between-school variation is accounted for in the fixed effects model, so differences between schools in teacher quality, resources and curriculum do not bias the estimates, although within-school variation of these factors can be interesting channels of mechanisms for further research. The coefficients of the models with control variables and the corresponding model that also includes fixed effects are very similar, which means that the covariates incorporate the sorting effect between schools to a high extent. However, there could be still some unobserved confounding factors (for example non-cognitive skills) that systematically differ within schools, between elite and general students which can bias the results.

Second, I conduct several robustness tests to account for other types of biases: I apply a different functional form (conditional logit), filter the sample (for example with and without 8-year track, and cut the upper 10% “always-takers”), use different covariates (a more detailed university aspiration variable and additional parental motivation proxies), and conduct panel analysis with year fixed effects.

5.1. Instrumental variables approach

An OLS estimate can be interpreted as a causal relationship if the conditional independence assumption holds, also called the selection-on-observables assumption for regression models (Angrist and Pischke, 2009 p.52-59). The key assumption for a linear causal model is that selection is only on observables, which is true in the case of a randomized experiment, but always questionable if one uses observational data in social sciences, as no matter how many observable variables we control for there is always the possibility that students in elite programs are different in their unobservable characteristics (omitted variables bias, or self-selection on unobservables). Using instrumental variables is a way to isolate exogenous variation in T to uncover its effect on Y, which helps us to solve the problem of bias from measurement error (in our setting it could be that NABC test scores are noisy measures of ability) and omitted variables (for example parental motivation, non-cognitive skills). Exogeneity is crucial, it means that the instrument is independent of the potential outcomes (Békés and Kézi, 2021 p.612), in our case, it should be independent of students' university enrollment probability if they participate in an elite program or not. In other words, unrelated to probability of university enrollment due to any reason other than the participation is an elite secondary school program, so it affects university enrollment only through elite program participation. A shortcoming of the IV estimator is that it gives a good estimate of the effect of the endogenous causal variable (T) only among observations, whose elite program participation (T) is affected by the exogenous instrument (Békés and Kézdi, 2021 p.619), these students are called compliers. In case of imperfect compliance, the effect on always-takers (supposedly high-ability high-SES students, and middle-ability high-SES students whose parents are committed to getting their child into an elite program) and never-takers (it is hard to speculate who could that be) cannot be measured by IV.

To address causality and test the robustness of the OLS model I apply IV approach. We need a variable that explains the selection into elite programs well (strong first stage) but does not explain our outcome (entering into higher education) through any other channel than elite program participation (exclusion restriction). There is one candidate for such an exogenous variance: the distance of home from the nearest elite program in grade 6. I will be using the regional variation (distance) from home to the closest elite program as an instrument, similar to the one proposed by Card (1999) generally and used by Schiltz et al (2019) for Hungary. In this setting, the exclusion assumption of IV is not met, since the geographical dispersion of schools and universities might not be independent of each other, and socioeconomic factors influence housing options thereby both distance from the nearest elite program and from the university. If we look at the 6th grade sample, there is a strong correlation between distance from an elite program and distance from a university (p = 0,52***). To solve this problem, I will use the proposed IV conditional on socioeconomic background variables and distance to the university. I obtain data to proceed with the IV estimation from the GEO database. Distances were calculated between the center of the following two ZIP codes: students' homes and the closest school with elite program, such as home to the closest university. Distance is measured in kilometers by car, and distance from the closest elite program takes the value of zero if there is a school with an elite program at the same ZIP code as the student's home address. I replaced the distance from university with 0 if someone lives at Budapest.

Distance from school as an instrument can be problematic if it is not strictly exogenous. For instance, higher-status parents sort themselves into the vicinity of better schools, and there was also a higher political pressure for the establishment of elite programs in wealthier regions and cities. Although the map in Figure 3 shows the location of elite programs where we can see that school sites are well dispersed in the country, so this is an unlikely problem, but we still need to take it into account when interpreting the results. One might argue, that where the student lives is conditionally (ceteris paribus previous test scores, and family background) independent of their chances of being selected to an elite program. And the only reason why students that live closer to an elite program enter such a program with greater probability is that it decreases their costs (for example less travel, more information, and better knowledge on that specific school, and network). A relationship exists if there is a strong first stage regression.

5.1.1. First stage: effect of distance on elite program participation

We can show that living farther away from an elite program (ceteris paribus test scores, family background, and distance from university) has a negative effect on entering elite programs. The results of the first stage regression are presented in Table 21, separately for the 6th and 10th grade cohorts, regressions include the entire set of control variables, the same as in the previous models.

Table 21: First stage results

Y: Elite	Sample 6	Sample 10
Distance from elite program	-0,005***	-0,005***
	(0,0006)	(0,0006)
Distance from university	0,001**	0,0006
	(0,0005)	(0,0005)
Observations	15538	13802
R-Square	0,1168	0,1169
F-statistic	56,7	65,2

Significant at ∗∗∗ 1%, ∗∗ 5%, and ∗ 10% levels. Standard errors are in parentheses. Models include f.e. and controls, see baseline specification under Table 12. Additional information: controls also include distance from the student's home to the closest school with elite program, and to the closest university. Outcome variable is the treatment (1 if the student attends elite program and 0 otherwise). Effective F-statistic of Montiel Olea and Pflueger (2013) is reported.

The F-statistic is greater than 10, which shows that the IV is not weak. The interpretation of the coefficient -0,005 is the following: students that live one additional km away from the nearest elite program (6- or 8-year long, whichever is closer), have a 0,5 percentage points less chance of entering an elite program, holding other variables constant. I emphasize again, that this effect is conditional on all variables included in the regression. The coefficient of distance from the closest elite program is the same in both cohorts, which implies the reliability of the results. As in our data, 20% go to these elite programs (only students in academic high schools are in the data), this is an important, albeit not very strong effect. Distance from the closest university, however, is positively associated with elite program participation - which could be due to the fact that both elite programs and universities are more concentrated in cities -, in addition, the effect size is smaller than for distance from school, and imprecisely estimated.

5.1.2. Two-stage least squares IV estimator

In this paragraph, I will present the IV estimator for the effect of elite program participation on university enrollment in the 6^{th} and 10^{th} grade samples. Table 22 below shows the results of the estimations: the effect of elite programs on university enrollment in 6^{th} and 10^{th} grade cohorts. OLS shows that students in elite programs have about 16-18 percentage points higher chance of enrolling in a university (see Table 12: Empirical results: enrollment and). This drops to roughly 4 percentage points if we control for many observed characteristics. The story does not change if we include school fixed effects (between school selection is much larger than within school selection), and the most preferred OLS coefficient is around 3,5 percentage points. If we instrument the treatment with distance, the effect size increases substantially (significant on the 5% level). That is, students that are on the margin (i.e. they are affected by the distance from the nearest elite program, so either chose this track because they are close to it, or not chose it because they are far, while otherwise they would do the opposite) are 10,1-7,8 percentage points more likely to enroll to university if they go to an elite program.

Table 22: 2SLS results (enrollment)

University enrollment	Sample 6		Sample 10	
	OLS	IV	OLS	IV
Elite	0,035***	0,101	0,034***	0,078
	(0,012)	(0,095)	(0,009)	(0,085)
Observations	17722	15538	17514	13802
Adjusted R-Square	0,307	0,274	0,336	0,327

Significant at ∗∗∗ 1%, ∗∗ 5%, and ∗ 10% levels. Models include f.e. and controls, see baseline specification under Table 12. Additionally, IV models control for the distance from home to the closest university.

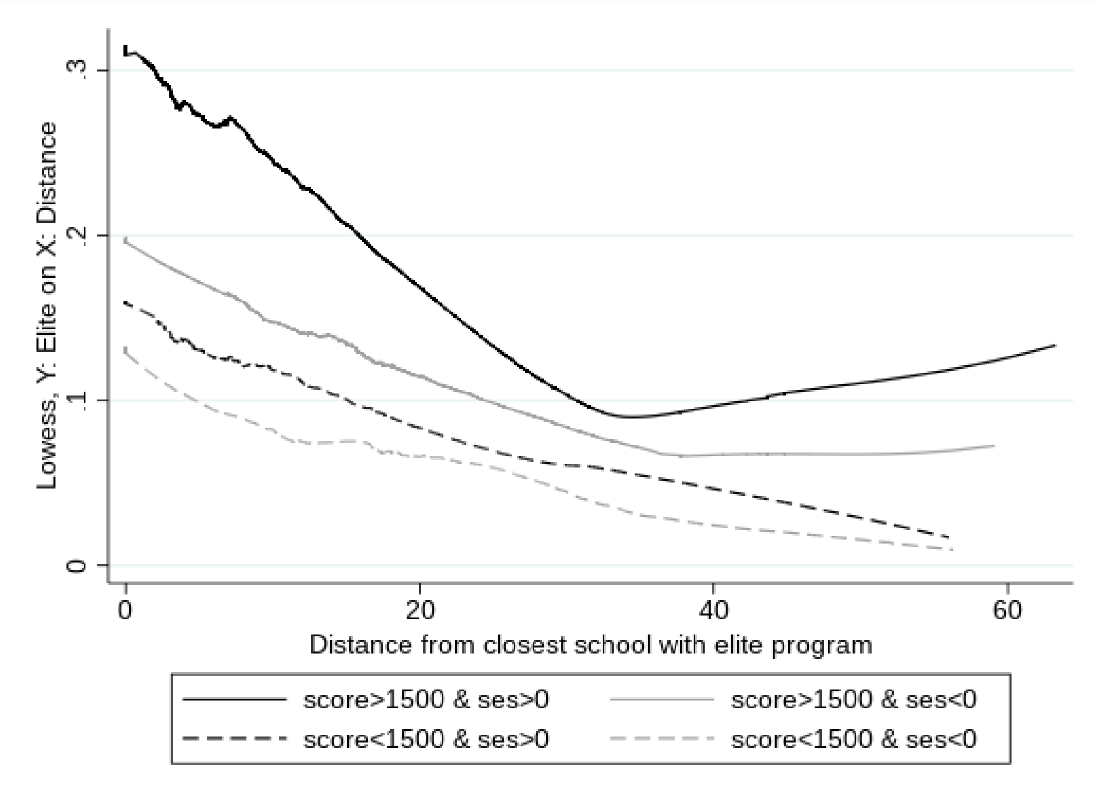
A possible reason why the conditional IV coefficient turned out to be insignificant is the substantial decrease in sample size due to the use of distance data: ZIP codes of students were missing in 2679 cases, those students immediately dropped out of the sample when merging the two datasets. If we consider the effect size, the roughly 8 percentage points effect – even if it is imprecisely estimated – is slightly higher than the corresponding OLS coefficient, so it supports the positive effect. However, if the exclusion restriction does not hold, the IV estimate can also be biased.

5.1.3. Interpreting the results – the role of heterogeneity and compliance

IV estimates should be interpreted carefully in the presence of heterogeneous effects for groups affected differently by the instrument, and imperfect compliance. Card (1999) concluded, that in models investigating the returns to education, IV estimates are roughly 20-40% above corresponding standard OLS estimates. The most plausible reason for this could be, that marginal effects are greater for subgroups - whose schooling decisions are mostly affected by the proposed IV -, than the average effect measured by OLS (Card, 1999 p.1855). Although the model is different, a similar mechanism holds in our setting. Hence, OLS effects of 3-4 percentage points, at the same time 8-10 percentage points in IV is not a surprising result.

He also highlighted, that in the presence of heterogeneous effects, IV might only "recover the weighted average of marginal returns for the affected subgroups" (Card, 1999 p. 1834), in our case students whose school choice was dependent on distance from school. Lowess graphs in Figure 8 show that the strongest association between distance and elite program is present in the case of high-ability, high-SES students (the black line includes students whose mathematics test score is above 1500 points and whose SES index is higher than 0), it means that they seem to be the subgroup that is affected the most by distance. High-ability, low-SES students are also strongly affected, but to a lesser extent, whereas distance is a less important factor for lower-ability students. Figure 8 also shows, that the closer the student lives to an elite program, the more distance is associated with elite program participation, and the relationship disappears around 35km, which means that the marginal effect of additional distance above this threshold is close to zero.

Figure 8: Lowess graph for the effect of distance on elite program participation



Although it is very hard to distinguish here, who the compliers are. That is, we do not exactly know, who those are, who modify their action due to variance in the exogenous variable (for example, those who live so far away from elite programs that they rather choose the alternative).

Another possible IV in this line of research could be the month of birth (see applications of a similar instrument at Black et al, 2011 and Szabó-Morvai et al, 2022), since there is a cut in school admission at 31. May meaning that all children who reached their sixth birthday by the end of May have to enroll in school in the upcoming September, but children who reach their sixth birthday after May can wait until next September. This results in better school achievement for students born in June, July, and August and worse for March, April, and May because they are similar aged students, but the first group (after the cutoff date) enroll later and have higher cognitive skills by that time, whereas the second group (before the cutoff date) enroll less mature. Although the same way as the month of birth can affect the probability of getting into elite programs (due to higher ability), it can affect the probability of enrolling in university as well, so in this case, it would not be an appropriate IV.

5.2. Robustness

During the analysis, I applied imputation and used different samples (sample 6 and 10) and subsamples of them (for example, without 8-year long track in 4.7.1. Primary school quality, or nonmissing GPA and aspiration in 4.6. Mechanism) so I find it important to check thoroughly whether these steps could have altered the results or not. I have conducted several robustness checks to see whether the main coefficient (3,5%: OLS estimate of enrollment effect in the most preferred model) is sable throughout different specifications. Essentially, my assumption is that these modifications did not change the model and its findings substantially, so results are not driven or influenced by these steps and the coefficients in the different samples are comparable. What is more, if the coefficient is stable in different samples and different periods, it implies the generalizability of the findings.

First, I run conditional logit models to test whether the choice of functional form matters in the regression. Although coefficients are not directly comparable, conditional logit gave similar results as OLS (linear probability model) in terms of direction and significance of the effect, which implies that the linearity assumption in OLS and the possibility of predicted values outside of the 0-1 interval is not a major issue. I chose to present the results of OLS due to a more intuitive interpretation of the coefficients.

Second, I tested whether certain subgroups of the sample could drive the results. In the analyses, I treated the 8- and 6-year long elite programs together, but the question may arise whether these two types of programs affect students differently since one is longer than the other and selection also happens at an earlier age. When I filtered out the elite students of 8-year long tracks from the sample (see Table 20), the coefficient became 3,1 percentage points, which is not far from the original 3,5 percentage points coefficient, in addition, when I applied a dummy variable to control for the possible differences between 6-and 8-year long programs, it turned out to be insignificant, as well as the interaction term. From these results we can conclude that the effect is similar in 6- and 8-year long programs. I also reduced sample by cutting the upper part of the ability-distribution. I defined a group of so-called "always-takers" in Sample 6 and Sample 10 as well: those students who are in the best decile in their sample based on either math or reading scores will probably go to university anyway, so I dropped them from the samples. This resulted in slightly greater coefficients, which is in line with the fact that IV estimates (where only compliers are considered and no always-takers) are greater than OLS estimates.

The variable transformations to deal with missing values (see page 53) also did not change the main coefficient considerably: treatment effect in sample 6 was 3,4% on enrollment before transformations of GPA and university aspirations (N=13742), and 3,5% afterward (N=17722). This implies that the imputation did not alter the results.

I also conducted a panel analysis using not only those students who finished 6^{th} grade in 2008 but I also included those two additional cohorts who finished 6^{th} grade in 2009 and 2010 as well. I couldn't examine the university enrollment of later cohorts because data is only available until 2017, and those students who attended 6^{th} grade in 2010 can enroll the earliest in 2016. This way, the sample size nearly tripled which resulted in 51.505 observations. In the pooled OLS model with controls and school fixed effects, I also included year fixed effects, because average enrollment rates varied yearly. In this panel sample, students of elite programs are associated with 5,1 percentage points greater probability of enrolling in a university, ceteris paribus. The coefficient is significant and higher than in sample 6, so it is also conceivable that the effect is typically somewhat greater than what I identified.

Table 23: Robustness of elite coefficient under different specifications (sample 6, enrollment)

Model	Coefficient of Elite: β (s.e.) [N]		
Baseline (with f.e. and controls)	3,5***	(1,2)	[17722]
Conditional logit	0,26***	(0,07)	[17480]
Without 8-year elite program	3,1**	(1,5)	[16027]
Without 6-year elite program	3,7**	(1,7)	[15459]
Without "always-takers"	4,5***	(1,4)	[14824]
Before imputation	3,4**	(1,4)	[13742]
Panel	5,1***	(0,8)	[51505]

Significant at *** 1%, ** 5%, and * 10% levels. Models include f.e. and controls, see baseline specification under Table 12. Additional information: models from second row are named based on their one modification compared to the baseline. Conditional logit: log of odds ratio is reported[31].

In addition, I checked whether a more detailed classification of university aspirations may capture some of the unobserved motivations. When I used 4 categories distinguishing between the intended levels of higher education attainment (no university aspiration, BA, MA, PhD) instead of the binary variable, the main coefficients did not change substantially. Similarly, we can observe how often parents attended parent-teacher meetings and visited the school on other occasions, which could proxy parental motivation. However, including these measures also did not alter the results. Table 24 shows the main coefficients in the different models: in the first row the baseline OLS results are presented, followed by a model where the only modification compared to the baseline is that I applied 4 university aspiration categories instead of 2, and the last row shows the coefficients if I complement the baseline model with the two additional parental motivation proxy as control variables.

[31] Odds ratio is calculated from the reported coefficients in the following way: $e^{0,26}$=1,30.

Table 24: Robustness of elite coefficients when including further controls

	Y: Enrollment, Sample 6	Y: Enrollment, Sample 10	Y: Degree, Sample 10
Baseline (with f.e. and controls)	3,5***	3,4***	3,4***
Use detailed aspirations	3,3***	3,2***	3,3***
Use parental motivation	3,4***	3,6***	3,6***
N	17722	17514	17514

Significant at *** 1%, ** 5%, and * 10% levels. Models include f.e. and controls, see baseline specification under Table 12. Additional information: models in second and third row are named based on their one modification compared to the baseline.

Robustness checks show a relatively stable coefficient throughout different specifications and indicate no sign of error due to the choice of linear form and imputation, and it seems reasonable to handle 8-and 6-year long elite programs together since there is no significant difference in effect sizes. The results of the panel model in 2008-2010 suggest that the estimated elite coefficient might be even higher than what was measured in the 6th grade cohort.

A shortcoming of the database is that post-secondary enrollment is only observed at Hungarian universities, so we cannot observe if and whether a student is enrolled in a foreign university[32]. This discrepancy in the data results in biased estimates in the most preferred model (the one with control variables and fixed effects) only if students with similar backgrounds within schools between elite and general classes show different tendencies in admission to foreign universities, which is presumably unlikely. Although if there is any difference, it probably results in a downward-biased estimate of the elite effect: given that higher-ability students sort into elite programs (see 4.2. Selection into elite programs) and elite programs increase student performance more than general programs (see 4.3. Impact on test scores), they might also have a greater effect on studying abroad due to peer effects and increased competitiveness. Cook (2023) studies the emigration patterns among Hungarian students and shows that 2-4% leave the country immediately after completing secondary education. She found that elite school attendance increases the probability of emigration by 1-2 % points. If the university enrollment probability is similar between those students who stay in

[32] Although there is an ongoing research which tries to identify Hungarian students in the Admin3 database who enrolled at a foreing university, and with the help of these results I might correct the enrollment variable later.

Hungary and those who relocate, these results imply, that among elite students the “underestimation[33]” of university outcomes is greater. As a consequence, my estimates are rather a lower bound of the actual effect size, due to the mismeasurement of tertiary enrollment abroad.

There are no exact statistics available where we could see the tertiary enrollment abroad individually and link it to the school type data but based on UNESCO statistics in 2012 (the year when the 6th grade cohort entered higher education) the number of Hungarian students studying abroad was 2,06% of the total tertiary enrollment in Hungary, and similarly to the EU average, this proportion has been growing since 2006: that time it was 1,5% in case of Hungary and it has reached nearly 2,5% in 2013 (Our World in Data, 2022). World Bank Education Statistics (2022) shows that the number of outbound internationally mobile tertiary students studying abroad in Hungary was 7831 in 2010 and this number increased to 12868 by 2018. As the number and proportion of Hungarian students studying abroad increase year after year, this problem will be worth revisiting later.

6. Discussion

There is a consensus in the literature that tracking harms equality, but the results about its effectiveness are inconclusive, which leads to the questioning of their existence. Context dependence is essential, as countries have different education systems and cultures, and studies (even those using standard RDD methods) find mixed results under different circumstances. It is also necessary to distinguish between short- and long-term impacts: while most studies find no short-term effects on test scores (although there are exceptions, see Pop-Eleches and Urquiola [2013] and Horn [2013]), long-term effects on university enrollment are generally positive.

This study exploits rich administrative matched data for students and institutions in Hungary to explore how students are being selected into elite secondary school programs and how attending them affects their test scores and university outcomes. I also developed a university ranking indicator that captures international recognition

[33] Falsely assigning 0 value for university enrollment, because they attend a university abroad which we cannot observe.

and quality to obtain an outcome variable with greater variability than simple enrollment rates, and measure also the institutions' quality which is also essential for human capital accumulation. The probability of attending elite programs is influenced by both family background (SES) and test scores at a similar magnitude, whereas within test scores, differences in mathematics seem to be a stronger driver of elite program admission than reading. The selectivity of 8- and 6-year long elite programs do not differ substantially, so it is reasonable to consider them as one treated group. Short- and long-run educational effects of elite programs were also compared, and the effects were positive in both cases. Given the same pre-treatment characteristics, test scores and GPA of students in elite programs change more favorably than their counterparts in general programs. However, changes in university aspirations do not differ significantly. The main focus of this research was post-secondary education outcomes: OLS- and PSM-estimates show significant, 3-4 percentage points effects of elite secondary school programs on university enrollment and completion rates, although the differences are much lower than raw enrollment and completion differences, and the impact on completion is nearly twice the magnitude. Similarly, I have found greater effects on MA enrollment than BA enrollment, and also a significant positive impact on the probability of attending internationally recognized universities, which implies that it is worth looking deeper than simple enrollment rate since it shows heterogeneity in university quality and difference in effect sizes for different outcomes. In the case of choosing certain fields of majors (STEM, medical, arts, law and governance), I have found insignificant effects.

Roughly half of the impact of elite programs is due to more favorable changes in the school performance of elite students than comparable general students: improvements in test scores and lower decreases in GPA. The results are heterogeneous between groups created based on gender, school performance, SES, and primary school quality, where students from more disadvantaged backgrounds generally benefit more from participating in elite programs. I used the distance from home to the closest school with an elite program as IV to uncover causality, and the two-stage least squares coefficient shows an 8-10 percentage points effect, which is close to the OLS estimate but slightly greater than that, although insignificant. Robustness checks show that the elite coefficient is stable throughout different specifications, and the panel model implies that it might be even higher in other periods than in sample 6.

Regarding the size of the impact on enrollment, my finding is in line with evidence of a similar study in the UK: Clark (2010) shows between 2-5 p.p., whereas the odds ratio of 1,2 at Lu (2021) translates to roughly 12% percentage point difference in our case[34] and comparable to the most preferred model (with controls and school fixed effects) that shows 3 p.p. but most probably this coefficient is a lower bound of the one in Lu's (2021) study because school fixed effect is a stronger over-control for school-level bias than the mean KS2 attainment in the other paper so in our setting we underestimate the effect compared to hers. Students who clear the threshold in a US boarding school have a 5-6 percentage points higher likelihood to enroll in college (Shi, 2020) compared to the (very high) baseline probability of 92% within non-admits close to the cutoff. Beuermann and Jackson (2022) found that attending preferred selective schools in Barbados increases the probability of entering and completing post-secondary university preparation programs[35] by 2 percentage points, which translates to higher university attainment for women later but not for men. The direction and significance of the effect are similar, the magnitude is comparable – even though grammar schools in the UK, boarding schools in the US, and education systems in Barbados and Hungary differ in several ways - which suggests that these results may generalize to other countries and other settings as well. On the other hand, it is also important to emphasize context dependency and that studies often find null effects as well.

Effectiveness and equality are two important factors in the characterization of education systems. If we look at the effect size of the OLS estimates, elite programs have a positive, significant association with not only university enrollment and completion, but with the quality of a university and further studies at MA level as well, even after controlling for observable student and school characteristics. These findings imply that elite programs are an effective way of educating their students, as they have a higher value-added in terms of university outcomes on average than their counterparts: general academic secondary school programs. In this sense, they deliver on the explicitly stated promises to increase the higher education participation of their

[34] If the odds ratio would be 1,2 in my setting, then it would mean that elite students have 1,2 times greater probability to attend university, conditional on the control variables. Taken that the enrollment rate of general students is around 60% in Sample 6, and 1,2*60=72 wich is the comparable elite students enrollment, this odds ratio would results in 72-60=12 p.p. difference.

[35] The Caribbean Advanced Proficiency Examination (CAPE) is two-year tertiary-level program, equivalent to the British Advanced levels examinations. Student who aim to attend university will take the CAPE, since passing certain units is a common admission requirement for universities (Beuermann and Jackson, 2022 p. 730).

students. On the other hand, the findings are controversial if we look at the equality effects. As students with higher SES sort into elite programs and these tracks increase the probability of positive long-term education outcomes (enrollment, completion, and ranking of university) it results in growing inequality of opportunity. However, we can observe an equalizing effect among students admitted to elite programs, as students with disadvantaged backgrounds (lower SES, lower school performance) benefit more from this program than their privileged peers. It would be essential to increase access to and participation in these elite programs for students with more disadvantaged backgrounds to counterweight the inequality effect.

My findings are subject to several caveats. As previously stated, elite programs in Hungary are unique because they are selected earlier in time (after grades 4, and 6), and selection occurs within schools where the alternative is the general academic route (after grade 8). As a result of context dependence, my findings are not directly comparable to the standard elite school literature, in which selection occurs later, at a single point in time and between schools. Second, the unconfoundedness assumption might not hold because unobserved characteristics (such as parental motivation and non-cognitive skills) probably correlate positively with both the treatment and outcome variables. As a consequence, some of my results based on this assumption may be upward biased. Third, while IV estimation may resolve part of the unobserved selection, the exclusion restriction may not be met, and the estimation could also be problematic due to weak IV. In addition, it is not trivial to determine who are the compliers, the subsample of students where we can identify the effect.

Last but not least, I would like to mention some unanswered questions that could be the topic of further research. First, it would be interesting to look into what policy measures could help increase the participation rates of more disadvantaged students in elite programs and make the selection less background-independent and more ability-based. Increasing information among students with less privileged status (with lower parental education, for instance) and in rural areas, or a form of teachers' recommendation system could be worth considering. Arany János Talent Program tries to bridge the gap and help students from more disadvantaged backgrounds attend academic secondary schools that prepare them better for higher education. However, segregation within schools between classes still exists as a challenge. Second, it would be interesting to discover further channels of the mechanism and find what causes more favorable changes in GPA, mathematics, and reading test scores in elite programs compared to general programs. Is it due to better teachers sorting to higher-achieving elite classes, differences in curriculum, or peer effects? Third, in this study, I analyzed the post-secondary education outcomes but did not include other important potential long-term effects on labor market participation, wages, and well-being due to limited time coverage of the database. The new wave of the Admin3 database will include data not only until 2017 but 2021, which will make it possible to observe early labor market outcomes as well, not only for the "oldest" 10^{th} grade sample (those students who are 10^{th} grade in May 2008 have become 29-30 years old by 2021) but also for a few younger cohorts, since these types of studies make sense when the observed sample is around 25-30 years old age. On the other hand, this database will also make it possible to analyze the university completion rates for the 6^{th} grade sample, where before-treatment school performance is available, so we can identify the effect more precisely. In contrast, the current database allows us to conduct this analysis only on the 10^{th} grade sample where the school performance variables (test scores, GPA and aspiration) are already under treatment, so I speculate that the estimated coefficients of BA completion and MA enrollment are the lower bound of the actual effect sizes in this sample.

References

Abdulkadiroglu, A., Angrist, J., & Pathak, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1), 137–196. https://doi.org/http://www.econometricsociety.org/tocs.asp

Ammermueller, A., J.-S. Pischke (2009). Peer Effects in European Primary Schools: Evidence From the Progress in International Reading Literacy Study. *Journal of Labor Economics*, 27 (3), 315–348.

Angrist, J. D., K. Lang (2004). Does School Integration Generate Peer Effects? Evidence From Boston's Metco Program. *American Economic Review*, 94 (5), 1613–1634.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Economics Books, Princeton University Press, edition 1, number 8769.

ARWU (2022): ShanghaiRanking's Academic Ranking of World Universities Methodology 2022. Last accessed: 13. September 2022. http://www.shanghairanking.com/methodology/arwu/2022

Békés, G., Kézdi, G. (2021). Data Analysis for Business, Economics, and Policy. Cambridge University Press.

Bennett, D. C. (2001). Assessing Quality in Higher Education. *Liberal Education*, 87(2).

Berkowitz, D. and Hoekstra, M. (2011). Does High School Quality Matter? Evidence from Admissions Data. *Economics of Education Review*, 30(2): 280-288.

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

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, *93*(2), 455–467. https://doi.org/10.1162/REST_a_00081

Borghans, B. L., Diris, R., Smits, W., & de Vries, J. (2020). Should We Sort It Out Later? The Effect of Tracking Age on Long-Run Outcomes. *Economics of Education Review, 75*.

Brooks, R. (2005). Measuring University Quality. *Review of Higher Education*, 29(1). 1-21. o. https://doi.org/10.1353/rhe.2005.0061

Brunello, G., Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Econ. Policy* 22 (52), 781–861.

Çakir, M., Acartürk, C., Alaşehir, O., Çilingir, C. (2015) A comparative analysis of global and national university ranking systems. *Scientometrics*, 103(3). 813-848. o. https://doi.org/10.1007/s11192-015-1586-6

Card, D. (1999). Chapter 30 - The Causal Effect of Education on Earnings. Pp. 1801–63 in *Handbook of Labor Economics*. Vol. 3, edited by O. C. Ashenfelter and D. Card. Elsevier.

Chatzimichael, K., Kalaitzidakis, P., Tzouvelekas, V. (2017). Measuring the publishing productivity of economics departments in Europe. *Scientometrics*, 113. 889-908. o. https://doi.org/10.1007/s11192-017-2509-5

Chmielewski, A. K. (2014). An International Comparison of Achievement Inequality in Within- and Between-School Tracking Systems. *American Journal of Education*, Volume 120, Number 3.

Claassen, C. (2015). Measuring university quality. *Scientometrics*, 104(3). 793–807. o. https://doi.org/10.1007/s11192-015-1584-8

Clark, D. (2010). Selective Schools and Academic Achievement. *The B.E. Journal of Economic Analysis & Policy, 10*(1). https://doi.org/10.2202/1935-1682.1917

Clark, D. (2022). The quality of lower-track education: Evidence from Britain. Working Paper 30174, *National Bureau of Economic Research*.

Clark, D. and Del Bono, E. (2016). The long-run effects of attending an elite school: Evidence from the United Kingdom. *American Economic Journal: Applied Economics*, 8(1):150 – 176.

Cook, R. (2023). Exploring the true extent of brain drain in Hungary: an analysis of emigration patterns among students. MA thesis, University of Cambridge

Csató László (2013). Rangsorolás páros összehasonlításokkal. Kiegészítések a felvételizői preferencia-sorrendek módszertanához. *Közgazdasági Szemle*, LX. évf. 1333-1353. o.

Csató László (2016). Felsőoktatási rangsorok jelentkezői preferenciák alapján. *Közgazdasági Szemle*, LXIII. évf. 27-61. o. DOI 10.18414/KSZ.2016.1.27

Csató, L., & Tóth, C. (2020). University Rankings from the Revealed Preferences of the Applicants. *European Journal of Operational Research*, *286*(1), 309–320.

Csóka, I. and Horn, D. (2022). A kisgimnáziumok hatása az egyetemekre való bekerülés esélyére. In: Szabó-Morvai, Á; Lengyel, B (szerk.) *Munkaerőpiaci tükör, 2021: adminisztratív adatok a gyakorlatban.* Budapest, Magyarország: Közgazdaság- és Regionális Tudományi Kutatóközpont, ELKH (2022) 309 p. pp. 171-177., 7 p.

Csóka, I., Neszveda, G., Sebestyén, G. (2019) 'Tudományos teljesítmény mérése a magyar felsőoktatás gazdasági képzéseiben', *Közgazdasági Szemle,* 66 : 7-8 pp. 751-770. http://dx.doi.org/10.18414/KSZ.2019.7-8.751

Deming, D. J. (2011). Better Schools, Less Crime? *Quarterly Journal of Economics* 126(4):2063–115

Deming, D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School Choice, School Quality, and Postsecondary Attainment. *American Economic Review*, 104(3): 991–1013. http://dx.doi.org/10.1257/aer.104.3.991

Divényi, János and Sóvágó, Sándor (2021). Do Elite Schools Benefit Their Students? (January 3, 2021). Available at SSRN: https://ssrn.com/abstract=3759304 or http://dx.doi.org/10.2139/ssrn.3759304

Dobbie, W. and Fryer, R. G. Jr. (2011). Exam High Schools and Academic Achievement: Evidence from New York City. *National Bureau of Economic Research* (NBER) Working Paper 17286.

Duflo, E., Dupas, P. and Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5), pp. 1739–1774. doi: 10.1257/aer.101.5.1739.

Estrada, R. and Gignoux, J. (2017). Benefits to Elite Schools and the Expected Returns to Education: Evidence from Mexico City. *European Economic Review*, 95, pp. 168–194.

European Commission/EACEA/Eurydice (2020). *Equity in school education in Europe: Structures, policies and student performance*. Eurydice report. Luxembourg: Publications Office of the European Union.

Fábri György (2016). Az egyetem értéke - Felsőoktatási rangsorok és az egyetemi teljesítmény. ELTE Eötvös Kiadó, Budapest.

Falch, T., Iversen, J. M. V., Nyhus, O. H., Strom, B. (2022). Quality measures in higher education: Norwegian evidence. *Economics of Education Review* 87, 102235. https://doi.org/10.1016/j.econedurev.2022.102235

FELVI (2022). Elmúlt évek statisztikái (2001/Á-2022/Á) - Jelentkezők és felvettek száma intézményenként. Last accessed: 14. October 2022. https://www.felvi.hu/felveteli/ponthatarok_statisztikak/elmult_evek/!ElmultEvek/index.php/elmult_evek_statisztikai/intezmenyenkent?filters%5Bsta_ev%5D=2014%2F%C3%81

Gurzó, K., Horn, D. (2015). A korai iskolai szelekció hosszú távú hatása: egy közpolitikai kísérlet tanulságai. *Közgazdasági Szemle* 62 : 10 pp. 1070-1096. , 27 p.

Györfi, A. (2021). Estimating gender differences in returns to elite schools. Working Paper, Vienna Graduate School of Economics

Hanushek, E. A., J. F. Kain, J. M. Markman, S. G. Rivkin (2003). Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics*, 18 (5), 527–544. [141,175]

Hazelkorn, E. (2007). The impact of league tables and ranking systems on higher education decision making. *Higher Education Management and Policy*, 19(2). 1-24. o. https://doi.org/10.1787/hemp-v19-art12-en

Hermann Z., Horn D., Köllő J., Sebők A., Semjén A., Varga J. (2022). A kompetenciaeredmények hatása a munkaerőpiaci sikerességre. *Közgazdasági Szemle* 69 : 2 pp. 177-198. , 22 p.

Hermann Z. and Horváth H. (2022). Tanári eredményesség és tanár–diák összepárosítás az általános iskolákban. Empirikus mintázatok három magyarországi tankerület adatai alapján. *Közgazdasági Szemle* LXIX. évf., pp. 1377-1406.

Horn, D. (2010a). A kisgimnáziumok szerepe a szelekcióban. In: Kolosi, T; Tóth, IGy (szerk.) *Társadalmi riport 2010*. Budapest, Magyarország : TÁRKI (2010) 551 p. pp. 408-429. , 22 p. http://www.tarsadalomkutatas.hu/kkk.php?TPUBL-A-927/publikaciok/tpubl_a_927.pdf

Horn, D. (2010b). The political background of structural changes in the educational system of Hungary, 1985-1994. Central European University, Center for Policy Studies – Working paper series, 45 p.

Horn, D. (2013). Diverging performances: the detrimental effects of early educational selection on equality of opportunity in Hungary. *Research in Social Stratification and Mobility*, 32:25–43. Social mobility and inequality in the life course: Exploring the relevance of context. http://dx.doi.org/10.1016/j.rssm.2013.01.002.

Hoxby, C. (2000). Peer Effects in the Classroom: Learning From Gender and Race Variation. Working Paper 7867, *NBER*.

Hoxby, C., G. Weingarth (2006). Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. *Working Paper, Harvard University*. Available at http://www.hks.harvard.edu/inequality/Seminar/Papers/Hoxby06.pdf

Imbens, G. W. (2015). Matching Methods in Practice: Three Examples. *The Journal of Human Resources*. Spring vol. 50 no. 2 373-419. doi: 10.3368/jhr.50.2.373

Johnes, G. (1987). Research Performance Indications in the University Sector. *Higher Education Quarterly*, 42. 54–71. o. https://doi.org/10.1111/j.1468-2273.1987.tb02114.x

Johnes, J. (2018). University rankings: What do they really show? *Scientometrics*, 115(1). 585-606.o. https://doi.org/10.1007/s11192-018-2666-1

Lavy, V., O. Silva, F. Weinhardt (2012). The Good, The Bad, and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics*, 30 (2), 367–414.

Leiden Ranking (2022). Indicators. Last accessed: 13. September 2022. http://www.leidenranking.com/information/indicators

Lu, B. (2021). Does attending academically selective schools increase higher education participation rates? *Cambridge Journal of Education*, 51(4):467–489.

Malamud, O., & Pop-Eleches, C. (2011). School Tracking and Access to Higher Education among Disadvantaged Groups. *Journal of Public Economics*, *95*(11–12), 1538–1549.

Marginson, S. (2014). University rankings and social science. *European Journal of Education*, 49(1). 45–59. o. https://doi.org/10.1111/ejed.12061

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, 110(2) 967-990. o. https://doi.org/10.1007/s11192-016-2212-y

Montiel Olea, J. L., and C. E. Pflueger (2013). A robust test for weak instruments.

Journal of Business and Economic Statistics 31: 358–369

Mullis, I. V. S., Martin, M. O., Goh, S., & Cotter, K. (Eds.) (2016). *TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science.* Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/encyclopedia/. Chapter: Overview of Education System, Hungary. Last accessed: 22. September 2022 http://timssandpirls.bc.edu/timss2015/encyclopedia/countries/hungary/

OECD (2019). PISA 2018 Results - Combined Executive Summaries Volume I, II & III. Last accessed: 08. October 2022 https://www.oecd.org/pisa/Combined_Executive_Summaries_PISA_2018.pdf

OECD (2021). Education at a Glance 2021: OECD Indicators. *OECD Publishing*, Paris, https://doi.org/10.1787/b35a14e5-en.

Oktatási Hivatal (2010). Országos kompetenciamérés – technikai leírás. Last accessed: 03. January 2023. https://www.oktatas.hu/pub_bin/dload/kozoktatas/meresek/orszmer2012/OKM_Technikaileiras.pdf

Oktatási Hivatal (2022). Középfokú felvételi eljárás: Prezentációk, tanulmányok, statisztikák. Felvételi a középfokú iskolákban a 2007/2008. tanévben. Last edited: 31. May 2022, Last accessed: 01. October 2022. https://www.oktatas.hu/kozneveles/kozepfoku_felveteli_eljaras/prezentaciok_tanulmanyok

Ono, H. (2001). Who Goes to College? Features of Institutional Tracking in Japanese Higher Education. *American Journal of Education*, *109*(2), 161. https://doi.org/10.1086/444265

Our World in Data (2022). Share of students studying abroad, 2008 to 2013. Last accessed: 4. November 2022. https://ourworldindata.org/grapher/share-of-students-studying-abroad?tab=chart&time=2008..latest&country=HUN~OECD+Member+countries~European+Union

Paternoster, R., Brame, R., Mazerolle, P. and Piquero, A. (1998). Using the Correct Statistical Test for the Quality of Regression Coefficients. *Criminology*, 36, 859–866.

Pollitt, C. (1990). Measuring University Performance: Never Mind the Quality, Never Mind the Width? *Higher Education Quarterly*, 44(1). 60-81. o. https://doi.org/10.1111/j.1468-2273.1990.tb01525.x

Pop-Eleches, C. and Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *American Economic Review*, 103(4):1289 – 1324

Psacharopoulos, G. and Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. *Education Economics*, 26(5), pp. 445–458. doi: 10.1080/09645292.2018.1484426.

QS (2022). QS World University Rankings methodology: Using rankings to start your university search. Last accessed: 15. September 2022. https://www.topuniversities.com/qs-world-university-rankings/methodology

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, 6(1):253–272

Safón, V. (2013). What do global university rankings really measure? The search for the X factor and the X entity. *Scientometrics*, 97(2). 223-244. o. https://doi.org/10.1007/s11192-013-0986-8

Scheeren, L., Bol, T. (2022a). Gender inequality in educational performance over the school career: The role of tracking. *Research in Social Stratification and Mobility*, Volume 77, 100661, ISSN 0276-5624. https://doi.org/10.1016/j.rssm.2021.100661.

Scheeren, L. (2022b). The Differential Impact of Educational Tracking on SES Gaps in Educational Achievement for Boys and Girls. *European Sociological Review*, jcac012, https://doi.org/10.1093/esr/jcac012

Schiltz, F., Mazrekaj, D., Horn, D., & De Witte, K. (2019). Does It Matter When Your Smartest Peers Leave Your Class? Evidence from Hungary. *Labour Economics*, 59, 79–91.

Scimago (2022). SIR Ranking Methodology. Last accessed: 13. September 2022. http://www.scimagoir.com/methodology.php

SciVal (2022a). Overview of metrics for academic institutions in Hungary, 2011-2020. (Types of publications included: all. Self-citations included: yes.) Downloaded: 20. June 2022. https://www.scival.com/overview/institutions?uri=Country/348

SciVal (2022b). Benchmarking Field-Weighted Citation Impact and publication year 2010-2020 for academic institutions in Hungary. (Types of publications included: Articles, reviews, conference papers, books, and book chapters. Self-citations included: no.). Downloaded: 17. September 2022. https://www.scival.com/benchmarking/analyse

Sebők, A. (2019). A KRTK Adatbank Kapcsolt Államigazgatási Paneladatbázisa. *Közgazdasági Szemle*, 66 (11). pp. 1230-1236. ISSN 0023-4346

Shi, Y. (2020). Who benefits from selective education? Evidence from elite boarding school admissions. Economics of Education Review, 74:101907.

Szabó-Morvai, Á., Horn, D., Lovász, A., De Witte, K. (2022). Universal Preschool and Cognitive Skills – the Role of School Starting Age as a Moderating Factor. *Early Childhood Research Quarterly* (under revision)

Tachibana, C. (2017). New tools for measuring academic performance. *Science*. https://doi.org/10.1126/science.opms.r1700173

Telcs András, Kosztány Zsolt Tibor, Török Ádám [2013]: Hallgatói preferencia-sorrendek készítése az egyetemi jelentkezések alapján. *Közgazdasági Szemle*, LX. évf. 290-317. o.

Terrin, É., & Triventi, M. (2022). The Effect of School Tracking on Student Achievement and Inequality: A Meta-Analysis. *Review of Educational Research*, *0*(0). https://doi.org/10.3102/00346543221100850

THE (2022). World University Rankings 2016-2017 methodology. Last accessed: 15. September 2022. https://www.timeshighereducation.com/world-university-rankings/methodology-world-university-rankings-2016-2017

Toutkoushian, R.K., Porter, S.R., Danielson, C., Hollis P.R. (2003). Using Publications Counts to Measure an Institution's Research Productivity. *Research in Higher Education*, 44(2). 121-148. o. https://doi.org/10.1023/a:1022070227966

Török Ádám (2008). A mezőny és tükörképei - Megjegyzések a magyar felsőoktatási rangsorok hasznáról és korlátairól. *Közgazdasági Szemle*, LV. évf. 874–890. o.

U-Multirank (2022). Catalogue of Indicators. Last accessed: 15. September 2022. https://www.umultirank.org/#!/measures?trackType=about&sightMode=undefined&section=undefined&name=null

US News (2022). How U.S. News Calculated the Best Global Universities Rankings. Last accessed: 15. September 2022. https://www.usnews.com/education/best-global-universities/articles/methodology

van Elk, R., van der Steeg, M., & Webbink, D. (2011). Does the Timing of Tracking Affect Higher Education Completion? *Economics of Education Review*, 30(5), 1009–1021.

Varga Júlia (2009). A tanárok elosztása a különböző szociokulturális hátterű tanulókat tanító iskolák között. Megjelent: Fazekas Károly (szerk.): *Oktatás és foglalkoztatás*: KTI Könyvek, 12. sz. MTA Közgazdaságtudományi Intézet, Budapest, 65–82. o.

Wagner, K., Dymes, L., & Wiggan, G. (2017). Tracking Students through Life: A Critical Structural Analysis of Academic Tracking of Mexican Immigrant Students in the United States and Korean Immigrant Students in Japan. *Urban Review: Issues and Ideas in Public Education*, *49*(5), 875–894.

Webometrics (2022). Methodology. Last accessed: 15. September 2022. http://www.webometrics.info/en/Methodology

Witte, K., Hudrlikova, L. (2013). What about excellence in teaching? A benevolent ranking of universities. *Scientometrics*, 96(1). 337-364. o. https://doi.org/10.1007/s11192-013-0971-2

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. http://www.jstor.org/stable/j.ctt5hhcfr

World Bank Education Statistics (2022). Total outbound internationally mobile tertiary students studying abroad, all countries, both sexes (number), series code: UIS.OE.56.40510. Last accessed: 4. November 2022. https://databank.worldbank.org/indicator/UIS.OE.56.40510?id=c755d342&report_name=EdStats_Indicators_Report&populartype=series#