The impact of elite high schools on academic performance in Hungary*

János Divényi, Sándor Sóvágó

This draft: October 7, 2015

Abstract

Selective schools attract high-achievers whose academic performance stands out. Does this achievement gap stem purely from selection or selective schools have value-added as well? This paper addresses this question by studying the effect of elite high school attendance on short-term academic performance in Hungary. Hungarian elite high schools track students at age 10 and 12 and offer special curriculum in addition to the high peer quality. Our research design provides non-parametric bounds on the effect of elite schools for those who actually attend them. For the identification, we assume that more able students select into the program and that the program does not impair the performance of participants. In addition to that, we tighten the bounds using parental education and the number of books at home as monotone instruments. We find that at least half of the achievement gap in numeracy and literacy test scores between elite and non-elite school pupils stem from selection. Furthermore, we provide evidence that elite high schools boost academic performance by at least 0.08-0.13 standard deviations after two years of treatment, and this positive effect persists after four years as well.

Keywords: gifted and talented education, selective education, non-parametric bounds

*János Divényi: Central European University, Hungary (e-mail: divenyi_janos@phd.ceu.edu). Sándor Sóvágó: VU University Amsterdam, The Netherlands, Tinbergen Institute (e-mail: s.sovago@vu.nl). The authors thank Gábor Kézdi, Bas van der Klaauw and Erik Plug for their comments and suggestions.

1 Introduction

Education systems typically allocate students into schools based on student's or parental preferences. Inevitably, certain schools are more popular or considered to be better than others, which leads to oversubscription. Often, oversubscribed schools prefer more able students, hence they admit students above a certain threshold, which is defined over previous grades or admission test scores. This mechanism gives rise to *elite* schools that admit only top students and whose students outperform their *non-elite* peers. However, it remains unclear if elite schools actually boost performance or the superior performance of their students only reflects selectivity.

This paper addresses this question by examining the impact of elite high schools on students' short-term academic performance in Hungary. Elite high schools are selective academic programs that track talented students into separate classes at age 10 and 12. The evaluation draws on an administrative dataset, which includes the national test scores for more than 500,000 students and which is complemented by an extensive survey on the student's socio-economic characteristics.

Our identification strategy builds on Manski (1997) and Manski & Pepper (2000) and provides non-parametric bounds on the treatment effect. The research design critically relies on three assumptions. First, we assume that elite schools admit more able students. Second, elite school attendance does not impair test scores, at least for those who actually choose these schools. Third, we assume that academic achievement relates positively to parental education and number of books at home. Our first two assumptions are not particularly informative by themselves as they presume that the treatment effect is non-negative and smaller than the naive OLS comparison. However, combining them with the proposed *monotone instrumental variables* (parental education and the number of books at home) of the third assumption potentially tighten the bounds on the parameter of interest.

We find that attending an elite high school in Hungary boosts academic achievement. In particular, both the 8-years and 6-years long academic tracks increase mathematics test scores by at least 0.05-0.1 standard deviations after two years of treatment. We find that the lower bound on the effect for reading is somewhat larger, it amounts to 0.08-0.13 standard deviations. Even if raw differences between elite high school and primary school students are massive (0.75-0.81 std for mathematics and 0.88-0.89 std for reading), we provide evidence that selection is quite strong and the effect of the program cannot exceed 0.34-0.37 for maths and 0.38-0.4 for reading after two years. The effect of the program persists after 4 years of exposure: we find that our estimated bounds shift to the right. Attending the 8-years long academic track for four years improves math test scores by at least 0.08-0.14 and reading scores by at least 0.09-0.16 standard deviations. The maximum effect size is 0.4 and 0.42, respectively.

Our paper contributes to the literature studying the causal effect of selective/elite schools on short-run academic achievement.[1] In particular, Abdulkadiroglu et al. (2014) and Dobbie & Fryer (2014) document that selective exam schools in New York and Boston, where students are exposed to better peers and to fewer non-white classmates, do not improve mathematics and English test scores. Although in a very different setting, Lucas & Mbiti (2014) also do not find any evidence that elite highs schools impact academic achievement in Kenya. Clark (2010) finds that selective grammar schools have modest impact on test scores, though the estimates are quite imprecise. Pop-Eleches & Urquiola (2013) estimate that selective Romanian high schools raise (high-stake) Baccalaureate grades by 0.02-0.1 standard deviations.

Our paper relates to the literature on the effects of gifted and talented (GT) programs. Card & Giuliano (2014) evaluate a US GT program that tracks talented students (high IQ) with students who scored well on pre-treatment tests. By exploiting the admission cutoff the authors do not find any effect on reading and maths test scores for the talented students, whereas they show that the program improves the performance of the high-achievers. Bui et al. (2014) do not find any effect for a very similar middle school program in the US. More recently, Booij et al. (2015) look at a Dutch pull-out GT program and document large positive effects on test scores in the short-run.

Our paper aims a methodological contribution as well. Most of the above cited articles on selective elite schools or gifted/talented programs exploit the admission policy rule: supply side constraints create sharp discontinuities in the admission probability. Therefore, these studies use (fuzzy) regression discontinuity design (RDD) as their research design and identify the local average treatment effect (LATE). We do not debate the internal validity of these estimates, however, it is contentious whether one can generalize them from the marginal student to other segments of the treated population. Our proposed identification strategy complements the RDD method and gives a more complete picture of the effect size for the entire treated population.

The remainder of the paper proceeds as follows. Section 2 briefly outlines the institutional details of the Hungarian education system. Section 3 describes the data, presents descriptive statistics and discusses our set identification strategy. Section 4 presents our main findings. Finally, Section 5 concludes.

[1] Some of the studies that we discuss look at medium- (such as university enrollment and grades) and long term outcomes (labor market outcomes, fertility) as well. Since these outcomes are beyond the scope of our paper, we do not discuss the results in detail.

Figure 1: The flow of students between school types (grade 6-8-10 in 2010-2014)

2 Institutional details

The Hungarian education system starts tracking students after the 4$^{\text{th}}$ year of the primary school, at age 10, when students can choose to attend 8-year long academic tracks. Later, at age 12, they can opt for a 6-year long academic track as well. Those pupils who are not tracked into the academic tracks finish their primary education at age 14 and enter secondary education for 4 years. Secondary education consists of three different tracks: 4-year long academic track, a relatively academically oriented vocational track and a more practical vocational track. Figure 1 depicts the distribution of a cohort across different tracks.

We focus on the students of the 8-years and 6-years long academic tracks and ask the question if they perform better in the academic tracks compared to having stayed in a primary school. The academic tracks share many similarities with those programs that has been studied by the economics of education literature, therefore we will refer to them (interchangeably) as *elite schools* or *special tracks* in the remainder of the paper.

Application to the 8-years and 6-years long academic tracks is based on the decision of the parents/students. Since the academic tracks are typically oversubscribed, the schools organize admission exams where students have to solve numeracy and literacy tests. While all applicants solve the same test in the country, the admission rule is school-specific, the admission grade is a combination of primary school grades, the written admission test and the oral admission exam. Consequently, the academic tracks are highly selective. As Figure 2 illustrates, students who enter elite schools experience a large increase in their peer quality. On the other hand, they are typically one of the best of their previous primary school classes (see Figure 3 about the ranks). Therefore, we label the 8-years and 6-years academic tracks as *elite schools* or *elite high schools* (Pop-Eleches & Urquiola (2013), Abdulkadiroglu et al. (2014), Dobbie & Fryer (2014), Clark (2010)).

The elite high schools offer more than selectivity, actually they share many similarities with gifted/talented programs (Bui et al. (2014), Card & Giuliano (2014)). These special tracks aim to attract motivated and talented students who plan to continue their studies in higher education. The curriculum is quite diverse across the country, but a common theme is that the lessons are organized in smaller classes. The small classes are often tracked further according to the skills and interests of the pupils: it is common to have more intensive classes in mathematics or natural

Figure 2: Grade-6 test scores of old and new peers of students who started elite school thereafter

Notes: The comparison is restricted on the students who choose 6-years academic track. "Old" refers to their peers in primary school whereas "new" refers to their peers in 6-years elite track. For both types of peers, the distribution of grade-6 test scores are plotted. Overlaid Cohort 2010.

Figure 3: Distribution of students' ranks in grade-6 by whether the student proceeded in 6-years track

Notes: Rank is defined by the percentile of the student's grade-6 score in the distribution of test scores within the same grade of the same school. Cohort 2010.

sciences. Besides, the program mandates to study a second foreign language. The special tracks are typically organized by secondary grammar schools that host the 4-year long academic tracks as well. Therefore, pupils are taught by better and more experienced teachers (Varga (2009)).

3 Data and research design

3.1 Data and descriptives

For the analysis we use Hungarian administrative data from the National Assessment of Basic Competencies (NABC). The assessment is similar to PISA but extends to every student in 6th, 8th and 10th grade. The survey is conducted annually in the last week of May. It measures literacy and numeracy skills (there is no science part) together with some background variables.

The skill scores are based on standardized tests which are filled by students in schools during classes. Therefore, they cover all but actually absent students (see Table 3.1 for number of observations). From 2008 onwards the test scores are standardized in a way that they are comparable across years and cohorts. 6th grade students in 2008 are serving as benchmark having an average score of 1500 with a standard deviation of 200; each class and cohort have been measured to them. The performance of the same students in different grades could be linked based on a permanent student identifier.

Figure 4 shows the average test scores of students choosing different school paths for students who finished grade 6 in 2010. There are huge differences between students choosing different tracks: the worst-performing students end up in vocational schools whereas better-achieving stu-

	2008	2009	2010	2011	2012	2013	2014	Total
A. Grade 6								
Raw data	107,654	100,620	96,898	94,047	92,082	93,907	92,544	677,752
w/o missing test score	100,255	93,382	89,940	87,293	85,669	87,150	85,538	629,227
w/o missing MIVs	79,143	74,793	73,836	71,609	71,153	73,103	72,330	515,967
w/ imputed MIVs	93,632	87,460	84,715	79,776	78,742	73,103	72,330	569,758
B. Grade 8								
Raw data	108,194	104,230	104,266	96,843	92,966	89,913	87,542	683,954
w/o missing test score	100,117	96,108	96,212	89,005	85,245	81,919	80,065	628,671
w/o missing MIVs	75,295	73,785	76,875	70,777	68,278	66,636	65,651	497,297
w/ imputed MIVs	88,181	85,276	89,619	83,160	80,057	74,733	73,452	574,478

Table 3.1: Sample size, missing data, data imputation

Notes: This table summarizes the sample size and the number of missing values by cohorts (2008-2014) and by grades (6-8). 7.1% and 8.1% of the test scores are missing in grade 6 and 8, respectively. We drop these observations from our sample. Furthermore, 18% and 20.9%)of the MIVs are missing in grade 6 and 8, respectively. We impute 8.5% and 12.2%of the data. (Note, that we cannot impute missing MIVs for students in grade 6 after 2013 as they have no other observation yet.)

dents go to elite schools and the differences seem only to widen through the years.

The tests are complemented by an extensive background survey on the student's socio-economic characteristics which they fill out together with their parents. Our proposed *monotone instrumental variables* are built upon this part of information. The background survey is voluntary, the average completing rate is around 75 percent. The third rows of Table 3.1 show how many students we lose due to missing data on background survey in the two grades we use throughout the analysis.

As each student are measured three times during her life, it is possible to impute some of the missing data in the background survey by entries of the same data from different years. The last rows of Table 3.1 inform about the effectiveness of the imputation. Appendix C discusses the details and the potential concerns.

In the analysis we pool the yearly data for grade 6 and 8. Appendix D provides some descriptive graphs about year-to-year dynamics of the main variables. We look at the fact that no clear trend emerges as justification for the pooling.

Table 3.2 summarizes the key variables we use by grades and programs (pooled by year).

	Primary	Elite8	Elite6
A. Grade 6			
Average math score	1485	1643	
Average reading score	1477	1655	
Mother's education			
At most primary	18.63%	1.61%	
Vocational	28.08%	10.16%	
Secondary	30.38%	28.70%	
High	22.91%	59.53%	
Father's education			
At most primary	15.62%	1.80%	
Vocational	43.85%	20.58%	
Secondary	22.61%	26.40%	
High	17.92%	51.22%	
Number of books at home			
Less than 50	17.44%	2.05%	
Around 50	14.74%	4.51%	
Max 150	22.06%	13.46%	
Max 300	15.51%	15.32%	
Between 300-600	12.62%	18.87%	
Between 600-1000	9.74%	21.42%	
More than 1000	7.90%	24.38%	
Number of observations	546,123	23,635	
B. Grade 8			
Average math score	1597	1758	1746
Average reading score	1555	1728	1717
Mother's education			
At most primary	18.10%	1.80%	2.03%
Vocational	29.67%	10.95%	11.64%
Secondary	31.43%	29.31%	29.28%
High	20.81%	57.94%	57.05%
Father's education			
At most primary	14.50%	1.71%	1.94%
Vocational	46.41%	21.80%	22.87%
Secondary	23.03%	26.39%	25.37%
High	16.06%	50.10%	49.82%
Number of books at home			
Less than 50	15.80%	1.80%	2.05%
Around 50	14.74%	4.34%	4.29%
Max 150	22.84%	12.74%	12.96%
Max 300	16.36%	15.46%	15.75%
Between 300-600	12.90%	18.79%	19.06%
Between 600-1000	9.70%	21.90%	21.56%
More than 1000	7.66%	24.96%	24.33%
Number of observations	517,614	23,671	33,193

Table 3.2: Descriptive statistics, grade

Figure 4: Average test scores for students choosing different tracks

Notes: Only the four main paths of Figure 1 are plotted (with matching colors): "Elite8" – students in 8-years academic track through all grades, "Elite6" – students in primary school in grade 6 and in 6-years academic track in grades 8 and 10, "Normal" – students in primary school in grades 6 and 8 and secondary school in grade 10 , "Voc." – students in primary school in grades 6 and 8 and in vocational school in grade 10. 95% confidence bands are plotted along.

3.2 Identification

Our primary interest lies in identifying the effect of attending elite schools for those who actually participated in the program, thus we focus on the average treatment effect on the treated (ATET):

$$ATET = \mathbb{E}\left[Y_1 - Y_0 | d = 1\right],$$

where d refers to the treatment status, Y_0 and Y_1 denote the potential untreated and treated outcomes, respectively. We believe that ATET is the most policy relevant parameter, since it measures the effect on the target group of the program. The special academic tracks are designed for talented students, therefore assessing their effect on low-achievers is not particularly interesting.

The vast majority of existing research identifies the effect of elite schools for the marginal students, that is for those who are close to the admission cutoff (e.g. Abdulkadiroglu et al. (2014), Dobbie & Fryer (2014), Card & Giuliano (2014), Pop-Eleches & Urquiola (2013)). Hence, these studies answer the question whether it is worth scaling up or down the programs. However, it is reasonable to assume that the program effect is heterogeneous, thus estimates for the marginal students should be interpreted as local average treatment effects (LATE). As a consequence, LATE estimates are uninformative about the overall effect of the program. For instance, having zero effect on the marginal student does not necessarily means that students in the higher segment of the ability distribution do not benefit from the program. Therefore, the overall effect on the treated population has important information content for both policymakers and parents.

This paper proposes a set identification strategy to assess the effect of attending elite schools on academic achievement. Set identification, pioneered and popularized by Charles Manski (Manski (1997), Manski & Pepper (2000), Manski (2013)), does not attempt to obtain point estimates for the treatment effect. Instead, it relaxes the often too restrictive point identification assumptions, and gives intervals that plausibly cover the parameter of interest. The advantage of set identification is that the identifying assumptions are usually much milder and more plausible. However, prediction is weaker and sometimes the estimated bounds are too wide, hence uninformative about the effect. Set identification is particularly attractive in our setting as we are mostly interested in the lower bound of the effect of elite schools, given that the literature mainly finds zero effect.

The identification challenge lies in finding out what would have been the outcome of elite school students if they had not been attending elite high schools. To bound the (mean) counterfactual outcome we propose three assumptions:

Assumption 1 (Monotone treatment selection)**.**

$$\mathbb{E}\left[Y_0|d=0\right] \leq \mathbb{E}\left[Y_0|d=1\right]. \tag{MTS}$$

Monotone treatment selection (MTS) assumes that pupils who attend elite schools have weakly higher mean test scores than their non-participating peers (Manski & Pepper 2000).[2] Essentially, the MTS assumption follows from the selectivity of the elite schools and from their admission procedure. Elite schools attract students with excellent academic skills and base their admission decisions on earlier test scores. As the admission procedure discriminates between low- and high performing pupils, admitted students have better pre-treatment academic skills and eventually higher pre-treatment test scores. Consequently, the potential outcomes of the participants exceed the potential outcomes of the non-participants.

Assumption 2 (Monotone treatment response)**.**

$$\mathbb{E}\left[Y_0|d=1\right] \leq \mathbb{E}\left[Y_1|d=1\right]. \tag{MTR}$$

Monotone treatment response (MTR) presumes that attending an elite school does not deteriorate the test scores of those who actually participate in the program (on average).[3] That is, the test results of the elite school participants are at least as high as their outcomes would be in the normal track. The type of the outcome variables that we examine – numeracy and literacy skills – supports the MTR assumption. Participating in more advanced math classes or having been required to read more books do not deteriorate core numeracy and literacy competencies. As numeracy and literacy

[2]Since we are interested in the ATET parameter, we use the MTS assumption only for the untreated potential outcome (Y_0).

[3]Originally, Manski (1997) proposes a stronger version of the MTR assumption, namely $Y_{0i} \leq Y_{1i}$ for all i. We weaken this assumption for two reasons. First, the plausibility of the assumption is questionable for low-ability non-participants as it is not obvious that they would benefit from a selective program targeted to high-achievers. Second, the weaker version of the assumption is sufficient to bound the ATET parameter.

are essential for general academic achievement, the more advanced curriculum is unlikely to drain resources from these competencies.

A potential concern about the validity of the MTR assumption is that the effect is heterogeneous and the program impair the performance of certain students. In particular, it might well be the case that pupils just above the admission cutoff do not benefit or even suffer from the program. These marginal students suddenly become the worst in their class and struggle with the more intensive curriculum. We believe that this argument does not challenge our identification strategy, since regression discontinuity studies mentioned above do not document significant negative effects for the marginal students. Admittedly, any evidence suggesting that certain segments of the ability distribution suffer from the program weaken our strategy, but it is important to note that such evidence does not invalidate it, since we assume that the program does not hurt on average for the treated population. That is, even in the presence of negative local effects, our identifying assumption can still hold.[4]

The combination of the MTS and MTR assumptions bounds the counterfactual outcome ($\mathbb{E}\left[Y_0|d=1\right]$) by quantities that can be recovered from the data. The MTR implies that the ATET parameter is non-negative, whereas MTS restricts the effect to be at most as large as the naive OLS comparison:

$$0 \overbrace{\leq}^{\text{MTR}} ATET \overbrace{\leq}^{\text{MTS}} \mathbb{E}\left[Y_1|d=1\right] - \mathbb{E}\left[Y_0|d=0\right].$$

The combination of the MTS and the MTR assumptions is not particularly informative by itself. It states only that the effect is non-negative and that selection is positive. Therefore we extend the identification strategy by a third assumption.

Assumption 3 (Monotone instrumental variable)**.** The variable z is a *monotone instrumental variable*, if for the values z_0 and z_1 of z, such that $z_0 \leq z_1$, it follows that

$$\mathbb{E}\left[Y_d|z=z_0\right] \leq \mathbb{E}\left[Y_d|z=z_1\right], \quad d=0,1. \tag{MIV}$$

The monotone instrumental variable (MIV) assumption claims that there is a monotone instrumental variable z, such that the potential outcomes are weakly increasing in z (Manski & Pepper

[4]Another potential worry is top-coding. One can argue that treated students would have had test scores that are close to the maximum, therefore the test are unable to measure any improvement. We argue that this is not the case empirically, as we do not find any bunching at maximum test score values in the data.

2000).[5] We propose two monotone instrumental variables: parental education and the number of books that the family has at home (number of books hereafter). We argue that these variables are both positively related to the potential outcomes and thus they satisfy the MIV assumption. Arguably, the genes that pupils inherit and the environment that parents provides are critical for student's short-term academic performance. We look at parental education and the number of books jointly as they operate through similar mechanisms: First, both parental education and the number of books are positively related to the cognitive skills of the parents. As these cognitive skills are genetically transmitted to the kids, the kids will perform better in school and have higher test scores. Second, both MIVs relate to the intellectual environment that the pupils are exposed at home. More educated parents or those who find it important to read books pay more attention to the education of their kids, which could lead to higher core competencies. Finally, the MIVs relate positively to the financial well-being of the family, which translates into more resources devoted to education.

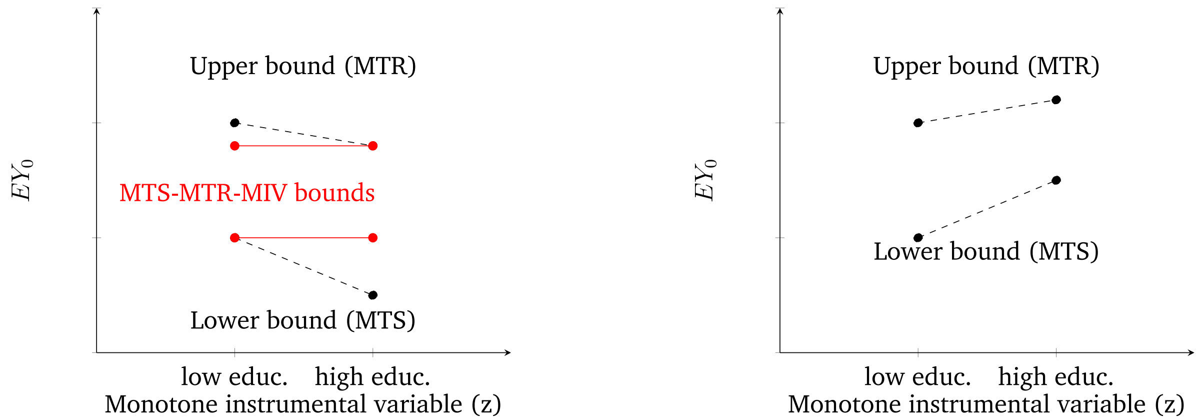
Figure 5 illustrates how the combination of the MTS, MTR and MIV assumptions can sharpen the bounds on the treatment effect.[6] For simplicity, suppose that monotone instrumental variable is parental education, which has two values (low and highly educated parents). We start by assuming that the MTS and MTR assumptions hold for each value of the monotone instrumental variable, which defines lower/upper bounds on the potential untreated outcome. This assumption is stronger than the original MTS and MTR assumptions, since they should hold for each subgroup defined by the monotone instrumental variable (Lafférs 2013). Now consider the case that the upper bound on the untreated outcome is higher for the low educated than for the highly educated parents (Figure 5a). Since the untreated potential outcome is monotone in parental education, the potential untreated outcome for the low educated parent group must be smaller than the upper bound for the highly educated parent category. Thus, the MIV assumption sharpens the upper bound by reducing the upper bound for the low educated category.[7] However, as Figure 5b shows, the MIV assumption does not necessarily tighten the bounds. If the MTS-MTR bounds are monotone in parental education, the MIV assumption does not have any impact on the bounds.[8]

Intuitively, the MIV assumption sharpens the MTS-MTR bounds if selection is strong: a large

[5]It is important to note that a monotone instrumental variables is not an *instrumental variable*. A variable is a valid instrument if MIV holds with equality, hence a monotone instrumental variable need to satisfy much milder conditions.

[6]Appendix A formally derives the MTS-MTR-MIV bounds.

[7]Similarly, assume that the lower bound on the untreated outcome is smaller for highly educated parents than for low educated parents. Since the untreated potential outcome is monotone in parental education, the potential untreated outcome for the highly educated parent must be larger than the lower bound for the low educated parent. Thus, the MIV



(a) Binding MIV

(b) Non-binding MIV

Figure 5: The mechanics of the MIV assumption

proportion of pupils with highly educated parents choose to go to elite school, such that even students with relatively low potential outcomes get admission. Furthermore, the elite schools admit only a small share of the students with low educated parents, but these students have very high potential outcomes. As a result, the observed outcomes in the low educated parent group exceed the observed outcomes in the highly educated parent category.[9] In other words, the observed mean outcomes (conditional on the treatment) should vary in the opposite direction with the MIV as what the assumption says about the potential outcomes (Richey (2013)).

The combination of two MIVs Although parental education and the number of books capture a similar idea, they are not perfectly correlated. Therefore we combine these MIVs to obtain even more informative bounds on the treatment effect. More formally, assume that t and z are two monotone instrumental variables. The MIV assumption for two instruments:

$$t_0 \leq t_1, z_0 \leq z_1 \Rightarrow \mathbb{E}\left[Y_d | t = t_0, z = z_0\right] \leq \mathbb{E}\left[Y_d | t = t_1, z = z_1\right], \quad d = 0, 1.$$

assumption sharpens the lower bound by increasing the lower bound for the highly educated category.

[8]More precisely, the bounds can still tighten due to selection: it might be the case that the distribution of the MIV in the overall population differs from the distribution in the population that we use for the estimation of the bound. For further details see Section 4 or Appendix A.

[9]For sharper lower bounds $\mathbb{E}\left[Y_0 | d = 0, z = z_1\right] \leq \mathbb{E}\left[Y_0 | d = 0, z = z_0\right]$ should hold, that is the average test score in the normal track declines with parental education. For sharper upper bounds (on the potential outcomes) $\mathbb{E}\left[Y | z = z_1\right] \leq \mathbb{E}\left[Y | z = z_0\right]$ should hold, that is the average test score in the population declines with parental education.

That is, students with weakly higher monotone instrumental variables have weakly larger potential outcomes (in expectation). To illustrate the assumption, consider the combination of the education of the mother and the number of books as MIVs. We assume that the potential outcomes weakly increase in the number of books for a given level of the mother's education. However, we do not presume any ordering between a low-educated family with many books and a highly-educated family with few books.

Estimation and inference We estimate the non-parametric bounds on the treatment effect by the corresponding sample means and empirical probabilities. Since the estimates exhibit finite-sample bias, stemming from taking minima and maxima of non-parametric estimates, we follow the bias-correction method by Kreider & Pepper (2007). Standard errors are based on 1,000 bootstrap replications and we construct the 95% confidence intervals following Imbens & Manski (2004). To ensure that our within MIV bin estimates are sufficiently precise, we require each bin to contain at least 1,000 observations.

4 Results

Our objective is to bound the average treatment effect on the treated (ATET). This section brings our identifying assumptions to the data and shows whether the proposed monotone instrumental variables can meaningfully narrow the bounds on the parameter of interest. First, we present the main results for academic achievement in grade 6 and then we turn to the estimates in grade 8. Finally, we discuss our findings.

4.1 The short-term effects of the 8-years long academic track

Table 4.1 presents our estimates for the effect of elite school attendance on mathematics and reading test scores in grade 6. More specifically, we bound how much better students in the 8-years long academic track performed compared to having had enrolled at a normal track. The naive comparison reveals that after being exposed to the treatment for 2 years, elite high school students outperform their non-elite school peers by 0.8 standard deviations in mathematics and by 0.88 standard deviations in reading (Panel A.). The massive gap between special and normal track test scores reflects the combination of selection and, potentially, the causal effect of the program. MTS presumes that selection is positive, hence the causal effect on maths and reading scores for the treated students cannot exceed 0.8 and 0.88 standard deviations, respectively. On the other hand, the MTR assumption ensures that the causal effect is at least zero. Importantly, the non-negative effect follows from our assumption and does not hinge on the data.

Panel B. of Table 4.1 combines the MTS, MTR and MIV assumptions using a single monotone instrumental variable. Parental education, measured separately by the father's and the mother's education, does not affect the lower bound on the treatment effect, thus we cannot exclude that elite high school enrollment does not impact test scores. Nevertheless, the upper bound on the ATET drops to 0.46-0.48 for maths and to 0.51-0.53 standard deviations for reading test scores.

To shed light on the mechanics of the MTS-MTR-MIV assumption, Figure 6c plots the MTS-MTR bounds on the mean untreated outcome for each parental education categories. Since selection to the program is not strong enough across the parental education categories, the monotony of the potential outcomes dominates and the MIV assumption does not tighten the bound (compare to Figure 5b). Consequently, the MTS-MTR-MIV and MTS-MTR lower bounds are equal. Seemingly, the situation is the same for the lower bound on the untreated outcome (upper bound on the ATET) as the bound increases monotonically (Figure 6c). However, the MTS-MTR-MIV upper bound drops by 0.3 standard deviations compared to the MTS-MTR bound for both test scores. The reason is that the lower bound on the untreated outcome is the weighted average of the parental education-specific lower bounds. These parental education-specific lower bounds are estimated on the untreated population, but they are weighted by the distribution of parental education based on the whole sample. As the children of highly educated parents are more likely to attend elite high schools (Figure 6a), we place more weight on the highly educated, which eventually increases the lower bound on the untreated outcome and pushes down the upper bound on the ATET.

A similar picture emerges when we look at the number of books variable: the lower bound remains zero, whereas the upper bound drops to 0.51 for maths and to 0.55 for reading. These bounds are quite comparable to those obtained by using parental education as the monotone instrumental variable (Figures 6c-6d).

Panel C. of Table 4.1 combines two monotone instrumental variables. The combination of parental education and the number of books variables further reduces the upper bound on the effect to 0.37-0.38 for maths and to 0.40-0.41 for reading. Admittedly, these upper bounds are still large, nevertheless, they decrease the MTS-MTR bounds by 50 percent and exclude unreasonably large treatment effects. Figure 7b depicts the lower bound on the expected untreated outcome for each number of books and mother's education categories. The figure illustrates that the observed math outcomes are monotonous in the education of the mother (for any given book category),

(a) Elite school attendance by mother's education

(b) Elite school attendance by the number of books

(c) Bounds on the untreated outcome: mother's education

(d) Bounds on the untreated outcome: the number of books

Figure 6: Elite school participation and MTS-MTR-MIV bounds in grade 6

Notes: Figures (a) and (b) display the probability of attending the 8-year long academic track in grade 6. In our sample 4.15 % of the students attended the 8-year long academic track between 2008 and 2014. Figure (c) and (d) show the MTS-MTR bounds on the mean untreated outcome for each values of the mother's education and the number of books, respectively. The red line indicates the upper bound, whereas the green line denotes the lower bound on $(\mathbb{E}[Y_0|z = z_i])$. The upper bound is based on the MTR assumption and calculated as the sample mean of the math test score within each bin. The lower bound is based on the MTS assumption and calculated as the sample mean of the math test score of the untreated students within each bin. The error bars denote the 95% confidence interval on the sample means.

however the non-monotonicity in the number of books for the low education bin makes the MIV binding and reduces the upper bound on the ATET.

More importantly, the combination of parental education and the number of books raises the lower bound on the treatment effect. We find that the casual effect of elite high school enrollment amounts to at least 0.06-0.1 standard deviations for mathematics and at least 0.08-0.13 standard deviations for reading. The positive lower bounds are explained by the fact that selection is so strong that the observed outcomes are no longer increasing in the number of books for lower parental education categories (i.e. for at most primary and vocational, Figure 7a). It is important to note again that in our setup the effect is non-negative by assumption. The positive lower bound is delivered by the combination of our identification assumptions and the data.

(a) Number of books and mother's education: lower bounds

(b) Number of books and mother's education: upper bounds

Figure 7: The combination of the number of books and mother's education: MTS-MTR-MIV bounds in grade 6

Notes: The colors refer to the education of the mother.

		Mathematics Lower/upper bound		Reading Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.803	0.000	0.883
	95% CI	0.000	0.813	0.000	0.892
B. Single monotone instrument (MTR - MTS - MIV)					
Father‘s education	Conventional	0.000	0.482	0.000	0.534
	Bias corr.	0.000	0.482	0.000	0.534
	95% CI	0.000	0.491	0.000	0.543
Mother‘s education	Conventional	0.000	0.464	0.000	0.516
	Bias corr.	0.000	0.464	0.000	0.516
	95% CI	0.000	0.474	0.000	0.525
Number of books	Conventional	0.000	0.505	0.000	0.547
	Bias corr.	0.000	0.505	0.000	0.547
	95% CI	0.000	0.514	0.000	0.556
C. Two monotone instruments (MTS - MTR - MIV)					
Father‘s education - number of books	Conventional	0.057	0.376	0.084	0.401
	Bias corr.	0.056	0.378	0.081	0.401
	95% CI	0.023	0.391	0.048	0.414
Mother‘s education - number of books	Conventional	0.108	0.370	0.138	0.406
	Bias corr.	0.099	0.372	0.131	0.408
	95% CI	0.052	0.384	0.080	0.419
# observations			569,758		

Notes: Panel A. presents the non-parametric bounds on the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds, the second row displays the bias-corrected estimates and the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

Table 4.1: The effect of elite school attendance on academic achievement in grade 6

4.2 Results in grade 8

When we look at the grade 8 test scores, we need to realize that the treatment group comprises of two subgroups: 8-years and 6-years long academic track students. Although the treatment is quite comparable for these groups, we examine these groups separately, since students in the 8-years long academic track are exposed to the treatment for a longer period.[10]

8-years long academic track Table 4.2 shows our main estimates for the 8-years long academic track. The upper bound on the ATET is 0.80 standard deviation for math and 0.89 for reading (Panel A.). These figures are almost identical to the MTS-MTR upper bounds in grade 6.

When we use parental education or the number of books as a single monotone instrumental variable, the upper bounds on the treatment effect drop to 0.49-0.53 for numeracy and to 0.55-0.57 for literacy (Table 4.2, Panel B.). In line with the MTS-MTR results, the upper bound for reading exceeds the upper bound for mathematics. Similarly to the grade 6 findings, a single monotone instrument does not affect the lower bound on the effect.

The combination of parental education and the number of books variable result in quite tight bounds, $[0.08, 0.40]$ and $[0.14, 0.39]$, for the maths scores using the father's and mother's education as the MIV, respectively (Table 4.2, Panel C.)). These bounds seem to shift to right compared to the bounds for grade 6 test scores. A possible explanation is that the negatively selected control group mechanically induces this shift. Another possibility is that the effect is actually larger than it was in grade 6. This explanation is not implausible as elite high school students are exposed to the treatment for 4 years by grade 8. We observe a very similar pattern for the reading test scores: the combination of the parental education and number of books MIVs narrow the bounds to $[0.09, 0.42]$ for the father's education and to $[0.16, 0.42]$ for the mother's education.

6-years long academic track Table 4.3 displays the effect of attending the 6-years long academic track on grade 8 test scores, that is the effect of elite high school attendance after 2 years exposure to the treatment. The MTS-MTR upper bound is 0.75 for mathematics and 0.83 for reading (Panel A.). These upper bounds are smaller than the upper bounds for the grade 6 test scores. Similarly to our previous estimates, a single monotone instrument reduces the upper bound on the effect by about 0.3 standard deviations, but leaves the lower bound unaffected (Panel B.). The combination of the parental education and number of books variable tightens the bounds further:

[10] Appendix B presents the results for the 8-years and 6-years long academic tracks together.

		Mathematics		Reading	
		Lower/upper bound		Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.807	0.000	0.886
	95% CI	0.000	0.817	0.000	0.895
B. Single monotone instrument (MTR - MTS - MIV)					
Father‘s education	Conventional	0.000	0.508	0.000	0.563
	Bias corr.	0.000	0.508	0.000	0.562
	95% CI	0.000	0.517	0.000	0.571
Mother‘s education	Conventional	0.000	0.492	0.000	0.548
	Bias corr.	0.000	0.492	0.000	0.548
	95% CI	0.000	0.501	0.000	0.556
Number of books	Conventional	0.000	0.529	0.000	0.567
	Bias corr.	0.000	0.528	0.000	0.567
	95% CI	0.000	0.538	0.000	0.575
C. Two monotone instruments (MTS - MTR - MIV)					
Father‘s education - number of books	Conventional	0.080	0.396	0.094	0.419
	Bias corr.	0.079	0.397	0.092	0.421
	95% CI	0.040	0.409	0.056	0.433
Mother‘s education - number of books	Conventional	0.154	0.390	0.162	0.421
	Bias corr.	0.144	0.393	0.155	0.423
	95% CI	0.089	0.405	0.106	0.435
# observations			541,285		

Notes: Panel A. presents the non-parametric bounds for the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds on the ATET parameter. The second row displays the bias-corrected estimates for the ATET parameter and finally the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

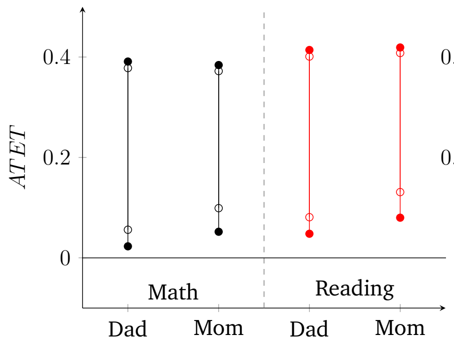
Table 4.2: The effect of 8-years academic track attendance on academic achievement in grade 8

it increases the lower bound into the positive domain (Panel C.). Consequently, the MTS-MTR-MIV bounds with two instruments allows us to exclude the possibility of zero effect, hence our result suggests that pupils who attend the 6-year long academic track benefit from the program already in the short-run. Based on the combination of parental education and the number of books, the estimated non-parametric bounds are comparable to the results for the short-term effect of the 8-years long academic track.

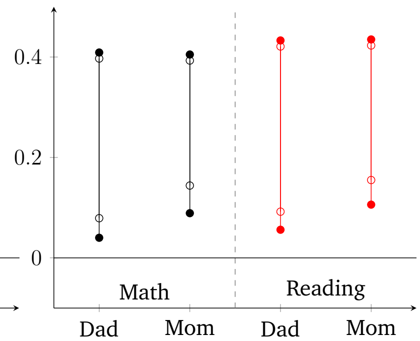
		Mathematics		Reading	
		Lower/upper bound		Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.745	0.000	0.827
	95% CI	0.000	0.753	0.000	0.835
B. Single monotone instrument (MTR - MTS - MIV)					
Father‘s education	Conventional	0.000	0.451	0.000	0.511
	Bias corr.	0.000	0.452	0.000	0.511
	95% CI	0.000	0.460	0.000	0.518
Mother‘s education	Conventional	0.000	0.437	0.000	0.497
	Bias corr.	0.000	0.437	0.000	0.496
	95% CI	0.000	0.445	0.000	0.504
Number of books	Conventional	0.000	0.471	0.000	0.513
	Bias corr.	0.000	0.471	0.000	0.513
	95% CI	0.000	0.480	0.000	0.520
C. Two monotone instruments (MTS - MTR - MIV)					
Father‘s education - number of books	Conventional	0.054	0.348	0.062	0.379
	Bias corr.	0.054	0.348	0.058	0.381
	95% CI	0.028	0.359	0.035	0.391
Mother‘s education - number of books	Conventional	0.114	0.343	0.123	0.381
	Bias corr.	0.108	0.345	0.121	0.382
	95% CI	0.067	0.355	0.087	0.392
# observations		550,807			

Notes: Panel A. presents the non-parametric bounds for the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds on the ATET parameter. The second row displays the bias-corrected estimates for the ATET parameter and finally the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

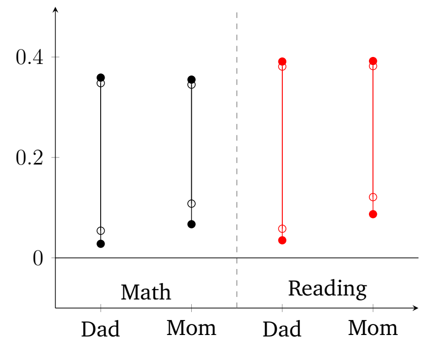
Table 4.3: The effect of 6-years academic track attendance on academic achievement in grade 8



(a) 8-years ac. track in grade 6



(b) 8-years ac. track in grade 8



(c) 6-years ac. track in grade 8

Figure 8: Summary of the main results

Notes: The figure summarizes the main results of the paper, that is the MTS-MTR-MIV bounds using parental education (dad or mum) and the number of books as monotone instruments. The circles correspond to the bias-corrected bounds on the ATET, whereas the dots depict the 95% confidence intervals. The black lines denote the estimates for maths, whereas the red lines refer to reading.

4.3 Discussion

Is this a big deal? One of the most relevant critique of our identification strategy is that it presumes a non-negative effect, hence, finding a small positive lower bound is not a big deal. More specifically, we assume that the effect is non-negative for each values of the monotone instrumental variable (pairs). One might worry that the program does actually hurt pupils who are coming from poor, low-educated families. Naturally, we cannot directly test this claim. However, the empirically testable implication of our MTS-MTR assumption, namely elite high school students outperform their primary school peers, holds for each MIV bin.

What do we add to the existing literature Our study is not the first that evaluates the performance of Hungarian elite high schools. Horn (2013) estimates the effect of both 8-years and 6-years long academic tracks on the value-added. The paper uses the distance to closest elite high school as an instrument and finds that elite high school attendance improves short-term academic outcomes. Even though we reach similar conclusions, we argue that the exclusion restriction for the distance to school instrument is violated. First, the potential outcomes of the kids are likely to affect the residential choice of the parents, and second, commuting distance has a direct impact on academic achievement. Therefore, we think that our identification strategy delivers more credible causal estimates and applies to a wider population.

Gurzó & Horn (2015) use a more credible instrumental variable which exploits the Hungarian schooling reform that led to the appearance of the elite schools. However, they look at long-term effects (employment and wage) rather than short-term ones (and find zero effect).

5 Concluding Remarks

Using a novel identification strategy in the literature, we have provided evidence that Hungarian elite high schools boost academic achievement in the short-run. It remains unclear, however, to what extent this result hinges on the Hungarian context and the identification assumptions. Nevertheless, the identification strategy that we propose does not depend on any institutional context and is applicable in other settings as well. Therefore, our methodology – with mild data requirements – can complement the prevailing RDD approach in the literature, hopefully yielding more nuanced picture about the effect of selective schools or gifted/talented programs.

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.
URL: *http://dx.doi.org/10.3982/ECTA10266*

Booij, A., Haan, F. & Plug, E. (2015), Enriching excellent students: Evidence from a gifted and talented program in secondary education. Manuscript.

Bui, S. A., Craig, S. G. & Imberman, S. A. (2014), 'Is gifted education a bright idea? assessing the impact of gifted and talented programs on students', *American Economic Journal: Economic Policy* **6**(3), 30–62.
URL: *http://www.aeaweb.org/articles.php?doi=10.1257/pol.6.3.30*

Card, D. & Giuliano, L. (2014), Does gifted education work? for which students?, Working Paper 20453, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w20453*

Clark, D. (2010), 'Selective schools and academic achievement', *The B.E. Journal of Economic Analysis and Policy* **10**(1), 1–40.
URL: *http://EconPapers.repec.org/RePEc:bpj:bejeap:v:10:y:2010:i:1:n:9*

Dobbie, W. & Fryer, Roland G., J. (2014), 'The impact of attending a school with high-achieving peers: Evidence from the new york city exam schools', *American Economic Journal: Applied Economics* **6**(3), 58–75.
URL: *http://www.aeaweb.org/articles.php?doi=10.1257/app.6.3.58*

Gurzó, K. & Horn, D. (2015), The long term effects of early educational selection. Manuscript.

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**(June), 25–43.
URL: *http://hdl.handle.net/1814/27817*

Imbens, G. W. & Manski, C. F. (2004), 'Confidence Intervals for Partially Identified Parameters', *Econometrica* **72**(6), 1845–1857.
URL: *http://dx.doi.org/10.1111/j.1468-0262.2004.00555.x*

Kreider, B. & Pepper, J. V. (2007), 'Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors', *Journal of the American Statistical Association* **102**, 432–441.
URL: *http://ideas.repec.org/a/bes/jnlasa/v102y2007mjunep432-441.html*

Lafférs, L. (2013), 'A note on bounding average treatment effects', *Economics Letters* **120**(3), 424 – 428.
URL: *http://www.sciencedirect.com/science/article/pii/S0165176513002668*

Lucas, A. M. & Mbiti, I. M. (2014), 'Effects of school quality on student achievement: Discontinuity evidence from kenya', *American Economic Journal: Applied Economics* **6**(3), 234–63.
URL: *http://www.aeaweb.org/articles.php?doi=10.1257/app.6.3.234*

Manski, C. F. (1997), 'Monotone Treatment Response', *Econometrica* **65**(6), 1311–1334.
URL: *http://ideas.repec.org/a/ecm/emetrp/v65y1997i6p1311-1334.html*

Manski, C. F. (2013), *Public Policy in an Uncertain World: Analysis and Decisions*, number 9780674066892 *in* 'Economics Books', Harvard University Press.
URL: *http://ideas.repec.org/b/hup/pbooks/9780674066892.html*

Manski, C. F. & Pepper, J. V. (2000), 'Monotone Instrumental Variables, with an Application to the Returns to Schooling', *Econometrica* **68**(4), 997–1012.
URL: *http://ideas.repec.org/a/ecm/emetrp/v68y2000i4p997-1012.html*

Pop-Eleches, C. & Urquiola, M. (2013), 'Going to a better school: Effects and behavioral responses', *American Economic Review* **103**(4), 1289–1324.
URL: *http://www.aeaweb.org/articles.php?doi=10.1257/aer.103.4.1289*

Richey, J. (2013), An odd couple: Monotone instrumental variables and binary treatments, MPRA Paper 56999, Munich Personal RePEc Archive.
URL: *https://mpra.ub.uni-muenchen.de/56999/*

Varga, J. (2009), A tanárok elosztása a különböző szociokulturális hátterű tanulókat tanító iskolák között (*Teacher sorting and pupils socio-economic characteristics*), *in* K. Fazekas, ed., 'Oktatás és foglalkoztatás', MTA Közgazdaságtudományi Intézet.
URL: *http://www.mtakti.hu/file/download/ktik/ktik12_09_elosztas.pdf*

Appendix A MTS-MTR-MIV bounds

This section derives the non-parametric bounds for the ATET parameter using the MTS, MTR and the MIV assumptions. The identification comprises of three steps:

Step 1 Bounds on $\mathbb{E}[Y_0]$

By the MTS and MTR assumptions for each values of the monotone instrumental variable z:

$$\mathbb{E}[Y_0|d=0,z=z_i] \leq \mathbb{E}[Y_0|d=1,z=z_i] \leq \mathbb{E}[Y_1|d=1,z=z_i], \quad i=0,1.$$

Now we reformulate the inequalities for the conditional expectation of the untreated outcome:

$$\begin{aligned}
&\mathbb{E}[Y_0|d=0,z=z_i]\cdot\mathbb{P}[d=1|z=z_i]+\mathbb{E}[Y_0|d=0,z=z_i]\cdot\mathbb{P}[d=0|z=z_i] \\
&\qquad\leq \mathbb{E}[Y_0|d=1,z=z_i]\cdot\mathbb{P}[d=1|z=z_i]+\mathbb{E}[Y_0|d=0,z=z_i]\cdot\mathbb{P}[d=0|z=z_i] \leq \\
&\qquad\quad \mathbb{E}[Y_1|d=1,z=z_i]\cdot\mathbb{P}[d=1|z=z_i]+\mathbb{E}[Y_0|d=0,z=z_i]\cdot\mathbb{P}[d=0|z=z_i], \\
&\qquad\quad \underbrace{\mathbb{E}[Y_0|d=0,z=z_i]}_{LB(0,z_i)} \leq \mathbb{E}[Y_0|z=z_i] \leq \underbrace{\mathbb{E}[Y|z=z_i]}_{UB(0,z_i)}, \quad i=0,1.
\end{aligned}$$

Note, that both bounds $(LB(0,z_i), UB(0,z_i))$ are empirically estimable quantities. The lower bound can be estimated by the mean of the untreated outcome for group z_i, whereas the upper bound is the mean outcome for z_i.
By the MIV assumption:

$$\begin{aligned}
LB(0,z_0) \leq \mathbb{E}[Y_0|z=z_0] &\leq \min\{UB(0,z_0), UB(0,z_1)\}, \\
\max\{LB(0,z_0), LB(0,z_1)\} \leq \mathbb{E}[Y_0|z=z_1] &\leq UB(0,z_1).
\end{aligned}$$

By the law of iterated expectations:

$$\begin{aligned}
&LB(0,z_0)\cdot\mathbb{P}[z=z_0]+\max\{LB(0,z_0),LB(0,z_1)\}\cdot\mathbb{P}[z=z_1] \\
&\leq \mathbb{E}[Y_0] \leq \\
&\min\{UB(0,z_0),UB(0,z_1)\}\cdot\mathbb{P}[z=z_0]+UB(0,z_1)\cdot\mathbb{P}[z=z_1].
\end{aligned}$$

Note that if selection is extreme, for example $LB(0,z_1) \leq LB(0,z_0)$ or $UB(0,z_1) \leq UB(0,z_0)$, the bounds will tighten. If that is not the case, then the upper bound on the expectation of the untreated outcome boils down to MTR bound, which equals to the population expectation of the outcome variable. However, the lower bound on $\mathbb{E}[Y_0]$ can tighten or widen without a *binding* MIV constraint. Even if $LB(0, z_0) \leq LB(0, z_1)$, we weight the conditional expectations ($\mathbb{E}[Y_0|d = 0, z = z_i]$) by the unconditional distribution of the monotone instrumental variable ($\mathbb{P}[z = z_i]$), hence the MTS-MTR-MIV lower bound will generally differ from the MTS-MTR lower bound.[11]

Using shorthand for the bounds on the expectation of the untreated outcome:

$$LB(0) \leq \mathbb{E}[Y_0] \leq UB(0).$$

Step 2 Bounds on $\mathbb{E}[Y_0|d = 1]$

Using empirically estimable quantities, we transform the bounds for the conditional expectation of the untreated outcome:

$$\frac{LB(0) - \mathbb{P}[d = 0] \cdot \mathbb{E}[Y_0|d = 0]}{\mathbb{P}[d = 1]} \leq \mathbb{E}[Y_0|d = 1] \leq \frac{UB(0) - \mathbb{P}[d = 0] \cdot \mathbb{E}[Y_0|d = 0]}{\mathbb{P}[d = 1]}.$$

Step 3 Bounds on the ATET parameter

$$\mathbb{E}[Y_1|d = 1] - \frac{UB(0) - \mathbb{P}[d = 0] \cdot \mathbb{E}[Y_0|d = 0]}{\mathbb{P}[d = 1]} \leq ATET \leq \mathbb{E}[Y_1|d = 1] - \frac{LB(0) - \mathbb{P}[d = 0] \cdot \mathbb{E}[Y_0|d = 0]}{\mathbb{P}[d = 1]}.$$

[11]Consequently, the MTS-MTR-MIV upper bound on the ATET parameter will be different from the MTS-MTR bound.

Appendix B Additional results

The main analysis studied the 8-years and 6-years long academic tracks separately. This appendix estimates the treatment effect for these track together in grade 8.

		Mathematics Lower/upper bound		Reading Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.771	0.000	0.852
	95% CI	0.0000	0.7775	0.0000	0.8578
B. Single monotone instrument (MTR - MTS - MIV)					
Father‘s education	Conventional	0.000	0.475	0.000	0.532
	Bias corr.	0.000	0.475	0.000	0.532
	95% CI	0.0000	0.4823	0.0000	0.5385
Mother‘s education	Conventional	0.000	0.460	0.000	0.518
	Bias corr.	0.000	0.460	0.000	0.518
	95% CI	0.0000	0.4668	0.0000	0.5239
Number of books	Conventional	0.000	0.495	0.000	0.535
	Bias corr.	0.000	0.496	0.000	0.536
	95% CI	0.000	0.5023	0.0000	0.5414
C. Two monotone instruments (MTS - MTR - MIV)					
Father‘s education -	Conventional	0.024	0.378	0.031	0.412
number of books	Bias corr.	0.023	0.379	0.028	0.413
	95% CI	0.0079	0.3865	0.0169	0.4201
Mother’s education -	Conventional	0.060	0.375	0.065	0.413
number of books	Bias corr.	0.055	0.376	0.062	0.414
	95% CI	0.0351	0.3838	0.0440	0.4209
# observations			574,478		

Notes: Panel A. presents the non-parametric bounds for the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds on the ATET parameter. The second row displays the bias-corrected estimates for the ATET parameter and finally the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

Table B.1: The effect of elite school attendance on academic achievement in grade 8

Appendix C Missing values

It is compulsory for the pupils to attend the competency test, hence, we observe the test scores for the vast majority of the students (92.9% in grade 6 and 91.9% in grade 8, Table 3.1).[12] Our MIVs, such as parental education or the number of books, come from the background survey that the students fill in on a voluntary basis at home (with their parents). Therefore, the non-response rate is significant, it amounts to the 18% and 21% of the sample in grade 6 and grade 8, respectively. Fortunately, students (potentially) attend NABC three times in their career, so they have multiple chances to fill in the background questionnaire. As we can link consecutive test results, we have the chance to impute the missing values by the past values. For imputation we start with the responses of the students in grade 10. For the missing values, we impute the values from grade 8, or if that is still missing, then we rely on grade 6 information.

While we believe that the parental education variable should be mainly constant between age 10 and 16 of the child, it might be the case that the number of books variable is trending over time. This appendix examines if such a trend is present and if so, what does it imply for our estimates.

Table C.1 displays the transition probabilities between grade 6 and grade 10 for the number of books. For instance, a student who reported to have around 50 books in grade 6, reports the same amount of books in grade 10 with the probability of 34%. She reports to have more books (max. 150) with 30% probability and less books (less than 50 books) with a 21% chance. The transition matrix suggests that the probability of stepping one bin up systematically exceeds the probability of stepping one bin down between grade 6 and grade 10. Overall, the chance that a student reports more books in grade 10 compared to grade 6 is 35.4%, whereas the probability of reporting less is 26.2%. Hence we conclude that the number of books variable is trending upward over time.[13] Next we examine how does this upward trend affect our results.

As a result of the upward trend, we misclassify certain individuals, that is we assign students to too low MIV bins. For now, we assume that missing values are random, i.e. they do not relate to treatment status or to potential outcomes, and that the trend is the same for treated and untreated pupils. It follows from the MIV assumption that imputation assigns high potential outcome students to the lower MIV group, therefore we overestimate $\mathbb{E}[Y|z=z_0]$ and $\mathbb{E}[Y_0|d=0,z=z_0]$. That implies that the MIV conditions are more likely to bind, hence we obtain too narrow bounds.

Since the response rate is higher among elite school students, we misclassify relatively more

[12] Students might be absent due to sickness or other personal reasons

[13] The same pattern prevails if we look at grade 6 - grade 8 or grade 8 - grade 10 transitions.

untreated students. Again, this misclassification will upward bias $\mathbb{E}\left[Y_0|d=0,z=z_0\right]$, which makes the upper bound on the ATET tighter. Similarly, $\mathbb{E}\left[Y_1|d=1,z=z_0\right]$ is higher compared to the no-missing-value case. However, there is another effect for the lower bound: the distribution of the treatment within the low MIV category shifts towards the untreated students, hence the effect is ambiguous on the lower bound:

$$\mathbb{E}\left[Y|z=z_0\right]=\overbrace{\mathbb{E}\left[Y_1|d=1,z=z_0\right]}^{\uparrow}+\overbrace{\mathbb{P}\left[d=0|z=z_0\right]}^{\uparrow}\underbrace{\left[\overbrace{\mathbb{E}\left[Y_0|d=0,z=z_0\right]}^{\uparrow}-\overbrace{\mathbb{E}\left[Y_1|d=1,z=z_0\right]}^{\uparrow}\right]}_{\leq 0}.$$

	Less than 50	Around 50	Max 150	Max 300	Bw 300-600	Bw 600-1000	More than 1000
Less than 50	0.52	0.28	0.14	0.03	0.01	0.01	0.01
Around 50	**0.21**	**0.34**	**0.3**	0.09	0.03	0.01	0.01
Max 150	0.06	**0.16**	**0.4**	**0.22**	0.1	0.05	0.02
Max 300	0.02	0.05	**0.24**	**0.31**	**0.22**	0.11	0.04
Bw 300-600	0.01	0.02	0.12	**0.21**	**0.31**	**0.23**	0.1
Bw 600-1000	0.01	0.01	0.05	0.1	**0.21**	**0.35**	**0.27**
More than 1000	0.01	0.01	0.03	0.05	0.09	0.22	0.6

Notes: The table is based on cohorts 2008-2010. Students in the sample filled in all the three surveys. The table displays the transition probabilities between between grade 6 and grade 10 for the number of books variable. The probability that a pupil reports more books in grade 10 than what she reported in grade 6 is 35.4%, whereas the probability that the student reports the same category is 38.4%. The treated students are 0.9 percentage point (with a standard error of 0.4) more likely to report higher number of books in grade 10 than in grade 6 (marginally significant).

Table C.1: Transition matrix: the number of books in grade 6 and in grade 10

We investigate if our results are sensitive to our imputation method. Table C.2 presents the main results for grade 6 without data imputation, in particular we focus on the number of books variable. We conclude that our results are not sensitive to data imputation, the combination of parental education and the number of books yield somewhat lower, but still positive lower bounds on the ATET. Similarly, the upper bounds are almost identical to our main estimates (Table 4.1).

		Mathematics		Reading	
		Lower/upper bound		Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.796	0.000	0.878
	95% CI	0.000	0.806	0.000	0.887
B. Single monotone instrument (MTR - MTS - MIV)					
Number of books	Conventional	0.000	0.500	0.000	0.545
		0.000	0.500	0.000	0.545
		0.000	0.510	0.000	0.554
C. Two monotone instruments (MTS - MTR - MIV)					
Father‘s education - number of books	Conventional	0.087	0.372	0.053	0.415
	Bias corr.	0.084	0.372	0.048	0.415
		0.040	0.385	0.017	0.427
Mother‘s education - number of books	Conventional	0.091	0.378	0.085	0.419
	Bias corr.	0.083	0.380	0.073	0.421
		0.029	0.392	0.023	0.432
# observations			515,967		

Notes: Panel A. presents the non-parametric bounds for the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds on the ATET parameter. The second row displays the bias-corrected estimates for the ATET parameter and finally the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

Table C.2: The effect of elite school attendance on academic achievement in grade 6: no data imputation

Appendix D Pooling of years

Throughout the analysis we pool the grade data across years. In this section of the appendix some descriptive figures are provided about the year-to-year dynamics of the main variables we use. We interpret them as exhibiting no clear pattern and thus allowing for pooling.

Figure 9: Average scores by year and grade

Notes: 95% confidence intervals are plotted as well but they are too narrow to see.

Figure 10: Distribution of mother's education by year and grade

Figure 11: Distribution of father's education by year and grade

Figure 12: Distribution of number of books by year and grade

Appendix E Placebo test

The way that students are tracked into 8-years and 6-years academic tracks offers a possibility to test the validity of the MIV assumption. Since we observe grade 6 test scores, we can bound the effect of attending the 6-years long academic track in grade 6. In grade 6 we do not expect to detect any treatment effect, since students start the 6-years long track 4 months after the competency test was taken.

What can we learn from the placebo test?

- We expect zero effect, since grade 6 test scores are pre-treatment outcomes
- We find positive lower bounds on the treatment effect
- Similarly to our main estimates, non-monotonicity in the high number of books and low parental education categories tightens the bounds

		Mathematics Lower/upper bound		Reading Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.867	0.000	0.891
	95% CI	0.000	0.875	0.000	0.898
B. Single monotone instrument (MTR - MTS - MIV)					
Father's education	Conventional	0.000	0.579	0.000	0.572
Mother's education	Conventional	0.000	0.562	0.000	0.558
Number of books	Conventional	0.000	0.598	0.000	0.580
C. Two monotone instruments (MTS - MTR - MIV)					
Father's education - number of books	Conventional	0.009	0.491	0.031	0.459
	Bias corr.				
Mother's education - number of books	Conventional	0.048	0.484	0.062	0.459
	Bias corr.				
# observations			390,445		

Notes: Panel A. presents the non-parametric bounds for the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds on the ATET parameter. The second row displays the bias-corrected estimates for the ATET parameter and finally the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

Table E.1: Placebo test: the effect of attending the 6-years long academic track in grade 6

- Consequently, our identifying assumption for two MIVs fails in this region
- Problem with the placebo test: we measure the test scores after the admission tests, so it is possible that there are anticipation/preparation effects, which we capture. We expect this effect much smaller than our treatment effect. Unfortunately, the bounds on the treatment effect are too wide to be able to say anything meaningful about the comparison

Appendix F Location type

This appendix collects the results for location type as a monotone instrumental variable.

		Mathematics		Reading	
		Lower/upper bound		Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.803	0.000	0.883
	95% CI	0.000	0.813	0.000	0.892
B. Single monotone instrument (MTR - MTS - MIV)					
Location type	Conventional	0.000	0.725	0.000	0.820
	Bias corr.	0.000	0.725	0.000	0.820
	95% CI	0.000	0.736	0.000	0.830
Location type (detailed)	Conventional	0.068	0.654	0.114	0.682
	Bias corr.	0.066	0.657	0.114	0.682
	95% CI	0.042	0.690	0.092	0.712
C. Two monotone instruments (MTS - MTR - MIV)					
Father‘s education - location type	Conventional	0.033	0.432	0.006	0.522
	Bias corr.	0.032	0.433	0.005	0.522
	95% CI	0.005	0.453	-0.007	0.537
Father‘s education - location type (detailed)	Conventional	0.212	0.252	0.104	0.402
	Bias corr.	0.219	0.250	0.085	0.415
	95% CI	0.178	0.290	0.035	0.456
Mother‘s education - location type	Conventional	0.099	0.360	0.008	0.498
	Bias corr.	0.093	0.360	0.001	0.499
	95% CI	0.059	0.390	-0.021	0.519
Mother‘s education - location type (detailed)	Conventional	0.075	0.290	0.218	0.281
	Bias corr.	0.088	0.269	0.212	0.285
	95% CI	0.012	0.318	0.171	0.330
Number of books - location type	Conventional	0.013	0.450	0.013	0.517
	Bias corr.	0.006	0.453	0.007	0.520
	95% CI	-0.012	0.473	-0.011	0.538
Number of books - location type (detailed)	Conventional	0.160	0.329	0.070	0.453
	Bias corr.	0.136	0.354	0.042	0.477
	95% CI	0.082	0.398	-0.000	0.515
# observations			476,126		

Notes: Panel A. presents the non-parametric bounds on the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds, the second row displays the bias-corrected estimates and the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

Table F.1: Additional results for location type - grade 6

..

		Mathematics Lower/upper bound		Reading Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.807	0.000	0.886
	95% CI	0.000	0.817	0.000	0.895
B. Single monotone instrument (MTR - MTS - MIV)					
Location type	Conventional	0.000	0.745	0.000	0.818
	Bias corr.	0.000	0.745	0.000	0.818
	95% CI	0.000	0.756	0.000	0.827
C. Two monotone instruments (MTS - MTR - MIV)					
Father's education -	Conventional	0.010	0.460	0.000	0.544
location type	Bias corr.	0.005	0.462	-0.000	0.546
	95% CI	-0.015	0.486	-0.002	0.561
Mother's education -	Conventional	0.098	0.376	0.000	0.532
location type	Bias corr.	0.098	0.377	0.000	0.535
	95% CI	0.061	0.410	-0.000	0.548
Number of books -	Conventional	0.030	0.450	0.017	0.509
location type	Bias corr.	0.028	0.453	0.014	0.512
	95% CI	0.010	0.475	-0.003	0.533
# observations		476,126			

Notes: Panel A. presents the non-parametric bounds for the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds on the ATET parameter. The second row displays the bias-corrected estimates for the ATET parameter and finally the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

Table F.2: Additional results for location type - grade 8 (8-years-long academic track)

Figure 13: Location type

		Mathematics		Reading	
		Lower/upper bound		Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.745	0.000	0.827
	95% CI	0.000	0.753	0.000	0.835
B. Single monotone instrument (MTR - MTS - MIV)					
Location type	Conventional	0.000	0.674	0.000	0.749
	Bias corr.	0.000	0.674	0.000	0.749
	95% CI	0.000	0.683	0.000	0.757
C. Two monotone instruments (MTS - MTR - MIV)					
Father's education -	conventional	0.005	0.407	0.000	0.489
location type	Bias corr.	0.001	0.408	0.000	0.491
	95% CI	-0.013	0.427	-0.001	0.502
Mother's education -	Conventional	0.065	0.340	0.000	0.475
location type	Bias corr.	0.065	0.341	0.000	0.477
	95% CI	0.039	0.366	-0.000	0.488
Number of books -	Conventional	0.023	0.401	0.010	0.459
location type	Bias corr.	0.022	0.403	0.008	0.461
	95% CI	0.009	0.420	-0.003	0.477
# observations		465,951			

Notes: Panel A. presents the non-parametric bounds for the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds on the ATET parameter. The second row displays the bias-corrected estimates for the ATET parameter and finally the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

Table F.3: Additional results for location type - grade 8 (6-years long academic track)

		Mathematics		Reading	
		Lower/upper bound		Lower/upper bound	
A. MTS - MTR					
	Conventional	0.000	0.771	0.000	0.852
	95% CI	0.0000	0.7775	0.0000	0.8578
B. Single monotone instrument (MTR - MTS - MIV)					
Location type	Conventional	0.000	0.704	0.000	0.778
	Bias corr.	0.000	0.704	0.000	0.778
	95% CI	0.000	0.711	0.000	0.784
Location type (detailed)	Conventional	0.031	0.688	0.056	0.718
	Bias corr.	0.030	0.693	0.056	0.719
	95% CI	0.020	0.709	0.047	0.732
C. Two monotone instruments (MTS - MTR - MIV)					
Number of books - location type	Conventional	0.023	0.401	0.010	0.459
	Bias corr.	0.022	0.403	0.008	0.461
	95% CI	0.009	0.420	-0.003	0.477
# observations			486,197		

Notes: Panel A. presents the non-parametric bounds for the ATET parameter under the MTS-MTR assumptions. The lower bound is zero by assumption. The 95% confidence interval is calculated using robust standard errors and it is based on Imbens & Manski (2004). In panels B. and C. three rows correspond to each monotone instrumental variable (or variable pair). The first row shows the non-parametric point estimates for the bounds on the ATET parameter. The second row displays the bias-corrected estimates for the ATET parameter and finally the third row shows the 95% confidence interval for the bias-corrected estimates. The bias corrected estimates and their confidence intervals are based on 1,000 bootstrap replications.

Table F.4: Additional results for location type - grade 8