

Tracking and specialization of high schools: heterogeneous effects of school choice

Olivier De Grootte* Koen Declercq†

October 1, 2020

Abstract

We analyze the impact of choosing an elite school on high school graduation in an early tracking system in Flanders (Belgium). While elite schools offer only an academic track, most other schools offer multiple tracks. On average, students experience a 3.3 percentage point increase in the likelihood of obtaining a degree. We find that the effects are heterogeneous. On average, students who self-select into elite schools do not experience an effect. However, students who do not choose an elite school would experience positive effects. Our results can be explained by different tracking decisions in both types of schools.

Keywords: elite schools, early tracking, marginal treatment effects

JEL: C31, I28

**Olivier De Grootte*: Toulouse School of Economics, University of Toulouse Capitole, email: olivier.de-groote@tse-fr.eu. Corresponding author at Toulouse School of Economics, 1 Esplanade de L'Université, 31080 Toulouse Cedex 06, France.

†*Koen Declercq*: CEREC, UCLouvain - Saint-Louis Bruxelles, email: koen.declercq@usaintlouis.be. 43 Boulevard du Jardin botanique, B-1000 Brussels, Belgium. We would like to thank the editor and three anonymous referees. We would also like to thank Peter Arcidiacono, Estelle Cantillon, Bart Cockx, Jan De Loecker, Kristof De Witte, Arnaud Maurel, Erwin Ooghe, Jo Van Biesebroeck, Frank Verboven, seminar participants at KU Leuven, Duke University and Toulouse School of Economics and participants of the 2015 LEER workshop on Efficiency in Education in Leuven, the 2016 Belgian Day for Labor Economists in Antwerp, the 2016 EALE conference in Ghent, the 2017 LEER Workshop on Education Economics in Leuven, the 2017 SOLE conference in Raleigh, the 2017 IWAE conference in Catanzaro, and the 2018 Matching in Practice Workshop in Cologne. Olivier De Grootte acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010. We would also like to thank the Flemish Ministry of Education for providing the dataset and the Research Foundation Flanders (FWO) and Fund for Scientific Research (F.R.S.-FNRS) for their financial support.

1 Introduction

Completing secondary education is important for opportunities in life. Among the 25-34-year-olds in OECD countries, the unemployment rate is 14.8% for individuals who did not complete secondary education, but only 7.8% and 5.8% for individuals who completed secondary and higher education, respectively. Furthermore, high school dropouts earn only 78% of the high school graduate wage. Nevertheless, 16% of young adults in OECD countries drop out without obtaining a degree (OECD, 2018).

We study the causal and heterogeneous impact of choosing an elite school on completing secondary education. We define an elite school as a school that offers only the academic track. We use a rich administrative dataset of all entrants in secondary education in Flanders, the Dutch-speaking part of Belgium. In Belgium, unemployment among high school dropouts is particularly high. Among the 25-34-year-olds, the unemployment rate for high school dropouts is 22.7% but only 9.0% and 4.7% for individuals who completed secondary and higher education, respectively (OECD, 2018).

Subsequently, we study how the organization of the tracking system can explain our results. Many countries apply some form of tracking and group students by their ability and preferences. Countries can be classified according to two different types of tracking systems. The first type of countries like Austria, Belgium, Germany, Greece, Japan and the Netherlands split students into academic and vocational tracks. The second type of countries like the US, Canada, UK and New Zealand apply course-by-course tracking and allow students to choose the level of their courses instead. While tracking into different programs often happens between schools, course-by-course tracking is applied within schools (Chmielewski, 2014).

In Flanders, students are tracked into different programs around the age of 13. The first year of secondary education consists of a comprehensive program. In the following years, students specialize in a certain program, within a track. Programs can be categorized into four hierarchically ordered tracks. The academic track prepares students for academic programs at university.¹ Other tracks prepare for professional programs in tertiary education or the labor market. As long as students perform well, they can choose from all of the programs. If they fail some courses, they may be excluded from certain options in the following year, or be required to repeat a grade.

There exist several types of schools in Flanders. While elite schools offer only the academic track, most non-elite schools are multi-track schools that offer programs in the aca-

¹Although there are almost no admission standards for higher education in Flanders, the tracks students follow in high school are important predictors of success at these institutions (Declercq and Verboven, 2015).

demic track together with programs in other tracks. Grouping students of different abilities into the same school is standard in countries that use course-by-course tracking, but it is less often the case in systems that split students early in academic and vocational tracks.² The supply of tracks within a specific school can affect study decisions. If schools offer programs in all tracks, tracking can occur within the same school. If schools specialize in one track, students may be less likely to choose the optimal track if this track is not available in the school. If students do not perform well, they can still switch to a lower track (downgrade) during their secondary education. Recent evidence shows that misallocated students can avoid the negative effects of early tracking by switching tracks when they are older (Dustmann et al., 2017). If more study options are available in the same school, misallocated students may be more likely to switch programs. Therefore, offering several tracks within a school can improve study outcomes by reducing switching costs. On the other hand, specialization in one track can have benefits in attracting a more homogeneous peer group or specialized teachers. It could also be more efficient if the school infrastructure is highly track-specific.

Enrollment in an elite school is a composite treatment. First, tracking decisions can be different in elite and non-elite schools due to a different supply of tracks. Second, elite schools attract students of higher ability and higher socioeconomic status leading to a different peer group composition. Finally, there could be differences in grading standards as schools have some autonomy in deciding grade progression. Notice that the Flemish government sets out common learning goals in each track that do not differ between elite and non-elite schools. Consequently, students in the academic track in elite and non-elite schools follow the same curriculum.

To study the treatment effects of elite schools on high school graduation, we face the following two problems: (1) endogeneity because of non-random self-selection into elite schools and (2) heterogeneity in the effects. We address the first problem by using distance as an instrument for school choice. The policy of free school choice and the large availability of different schools within a neighborhood make distance a reasonable instrument for school choice in Flanders, conditional upon controlling for neighborhood characteristics.

We allow for heterogeneous treatment effects in different ways. First, we follow the framework of Heckman and Vytlacil (2005) and estimate Marginal Treatment Effects (MTEs) to compute average treatment effects. The richness of our data allows us to use standard semi-parametric estimation techniques to recover the Average Treatment effect on the Treated

²Some other countries also have multi-track schools. See for example the “kooperative Gesamtschulen” in Germany (Mühlenweg and Puhani, 2010), or comprehensive schools in the Netherlands (Oosterbeek et al., 2020).

(ATT). We compare this estimate to an estimate that is derived from a potential outcome model that uses limiting functional form and distributional assumptions. This approach also gives us an estimate of the Average Treatment Effect (ATE) and the Average Treatment effect on the Non-Treated (ATNT) and illustrates how observable background characteristics can explain differences in treatment effects. Finally, we apply the factor structure in Aakvik et al. (2005) to identify distributional treatment effects: the percentage of students that would experience a positive or negative treatment effect from elite schools.

We obtain the following main findings. On average, students are more likely to obtain a high school degree if they start at an elite school. We obtain a significantly positive ATE of 0.033, which implies that choosing an elite school, on average, increases the probability of graduating from high school by 3.3 percentage points. We also find important heterogeneity in the treatment effect, reflected in a small and statistically insignificant ATT and a large ATNT of 5.1 percentage points. We conclude that students who do not choose elite schools would benefit most from treatment. Next, we find that the average treatment effects still mask a lot of heterogeneity. We find more positive effects for students who had repeated a grade before starting high school and students with a low-educated mother. Furthermore, our results indicate that there is heterogeneity within both groups of treated and non-treated students, with a substantial fraction gaining and suffering from elite schools in both groups.

To study how the tracking of students is influenced by the type of school and whether this explains the heterogeneity we find, we extend the potential outcomes framework by modeling the track choice in the second year of high school. We find that choosing the academic track raises the probability of graduating from high school. Moreover, elite schools encourage students to study the academic track, but primarily the non-treated students. This is because most treated students would choose an academic track in any school. At the same time, treated students are very reluctant to leave an elite school and aim for a lower track when they do not perform well. Consequently, this puts them at risk of dropping out.

Our findings have important policy implications. Encouraging students to start at elite schools can raise the overall degree completion in secondary education. However, there is a lot of heterogeneity in the effect of treatment within both the group of treated and non-treated students and the reverse selection on gains shows that students are not optimally choosing their school. An optimal policy would, therefore, allocate students to the type of school that benefits them most.

Our paper contributes to three strands of literature. First, we contribute to the literature on elite schools. Similar to our study, Clark and Del Bono (2016), Dustmann et al. (2017), Guyon et al. (2012) and Oosterbeek et al. (2020) analyze the effects of attending a school that specializes in academic programs. Pop-Eleches and Urquiola (2013) and Abdulkadiroğlu

et al. (2014) also study the effects of elite schools but define them by the quality of peers. In contrast to previous studies, we study the observed and unobserved heterogeneity in the treatment effect. We characterize the degree of heterogeneity by comparing the ATT and ATNT, but also the percentage of students who gain and lose. This shows that averages can mask large differences between students. Moreover, identification strategies in other papers identify an average treatment effect for specific groups in the population. Our results show that the large degree of heterogeneity in treatment effects makes it difficult to extrapolate these results to the total population.³ We also show how the supply of tracks in each type of school can explain heterogeneity in the effects.

Second, we contribute to the literature on educational tracking. Several papers have looked at the consequences of tracking students at an early age (Fu and Mehta (2018), Hanushek and Woessmann (2006), Pekkarinen et al. (2009), Roller and Steinberg (2020)).⁴ Other characteristics of tracking systems remain largely unexplored, except for Cockx et al. (2018), De Groote (2020) and Dustmann et al. (2017) who investigate the impact of switching tracks at a later age. We contribute to this literature by comparing the impact of tracking between and within schools.

Finally, we apply advances in the literature on estimating treatment effects proposed by Heckman and Vytlacil (2005) and Aakvik et al. (2005). We can identify semi-parametric MTE estimates to derive the ATT and compare it with the results of a model that estimates a potential outcomes model with functional form and distributional assumptions.⁵ Imposing these assumptions also gives us additional insights into the degree of heterogeneity of the treatment effect. We extend the potential outcomes model with the track choice in high school. Simultaneously modeling both choices allows us to study how school choice influences track choice and provides insight into the underlying mechanism that explains the effects of elite school attendance.

The paper is organized as follows. We start by providing an overview of secondary education in Flanders and introduce our dataset. Next, we discuss the empirical framework in

³Oosterbeek et al. (2020) study observed heterogeneity in the treatment effect based on test scores before the start of high school. Their study considers only students who were qualified to enter the highest track and decided to apply for an elite school. We look at decisions before tracking occurs and we characterize both observed and unobserved heterogeneity in the treatment effect.

⁴See also Cummings (2017) and Duflo et al. (2011) for evidence in developing countries.

⁵MTEs are usually only identified for a subset of individuals, such that it is not possible to estimate the ATE, ATT or ATNT nonparametrically. See, for example, Doyle (2007 and 2008) and Galasso and Schankerman (2015). Other papers rely on parametric assumptions on the shape of the MTE curve or the underlying behavioral model to improve common support or obtain sufficiently precise estimates (Carneiro et al. (2011), Carneiro et al. (2016), Cornelissen et al. (2018), Basu et al. (2007), and Nybom (2017)).

Section 3 and discuss the validity of our instrument in Section 4. Section 5 discusses the main results, Section 6 the underlying mechanism and Section 7 concludes. Online Supporting Information Appendices A-D provide details about the construction of the dataset, sensitivity analyses, a discussion on the role of control variables and additional tables and figures.

2 Secondary education in Flanders

In this section, we discuss the institutional context in Flanders and show descriptive statistics about enrollment and study outcomes for students choosing elite and non-elite schools.

2.1 Institutional overview

Flanders is the Dutch-speaking part of Belgium, located in the North. It consists of approximately 60% of the population of 11 million inhabitants. After finishing primary school, students enroll in secondary education, usually at the age of 12. Students can choose between all schools in Flanders since school choice is not geographically restricted and free school choice is law-enforced.⁶ In the first year of secondary education (grade 7), almost all students start in a comprehensive program. A small proportion starts in a vocational-preparatory program. As these students often did not successfully finish primary education, we do not consider this group in our analysis.

After the first year, students can choose between programs in four hierarchically ordered tracks.⁷ The academic track is the highest and prepares students for academic programs in higher education (mostly four- or five-year programs). The technical track and the artistic track prepare students for professional programs in higher education (mostly three-year programs) or the labor market. Students can also choose the vocational track, which prepares them for the labor market. Both students' preferences and scores in the first year determine the track. If they perform well, they can choose between all tracks. Students who fail on some courses may be excluded from certain programs or tracks. If they fail on too many important courses, they can also be required to repeat a grade. While mobility between tracks is generally possible, we observe almost only downward mobility. Note that this

⁶Capacity constraints were uncommon, but we cannot exclude that some schools were oversubscribed. The Belgian constitution, however, protects free school choice and prevents schools from cream skimming. If the school is capacity constrained, it must add students to a waiting list, and if places become available, it must respect the order of this list.

⁷Officially the distinction between tracks exists only from the third year on (grade 9). However, at the start of the second year, students decide on elective courses that prepare them for a particular track. We therefore classify tracks from the second year on.

mobility is not always due to the restrictions of continuing the same track but often results from a choice by students or their parents.

For each study program in each grade, the Flemish government sets out common learning goals, and schools organize their curriculum accordingly. The supply of these programs differs considerably between schools. According to the tracks offered at schools, we distinguish between two types. The first type of school is an elite school. They offer only programs in the academic track. The second type of school is a non-elite school. Most non-elite schools are multi-track schools and offer programs in the academic track together with programs in other tracks. However, some non-elite schools offer only programs in the technical, artistic and/or vocational track.⁸

2.2 A first look at the data

We use a rich dataset provided by the Flemish Ministry of Education. We observe all students who started in the comprehensive program in secondary education in 2003 and 2004.⁹ We follow these students in secondary education and observe them for six years if they graduate within the minimum duration of the program or nine years at most if they graduate after some delay. Every school year, we observe their school and program. We also observe when they obtain a high school degree and their study delay.

Table 1 summarizes the characteristics of students and their neighborhood in the two types of schools. 26.7% of all students start at an elite school. 48.6% of the students starting at elite schools are boys, while this is 49.9% at non-elite schools. Students who repeated at least one grade before secondary education are less likely to choose an elite school. Grade repetition can be interpreted as a proxy for ability because students who do not meet the minimum standards can be required to repeat a grade during kindergarten or primary education. Next, disadvantaged students are less likely to start at elite schools. The most notable effect is for the educational level of the mother. In elite schools, 55.8% of students have a mother with a higher education degree, while this is only 32.4% in non-elite schools.

The following panel shows neighborhood characteristics at the level of the municipality and the level of the statistical sector. In Belgium, each municipality is divided into several statistical sectors that split each municipality into homogeneous neighborhoods.¹⁰ Aggregate

⁸Our results do not change when we exclude the second type of non-elite schools from the analysis (see Supporting Information Appendix B).

⁹Supporting Information Appendix A provides more details about the data selection process.

¹⁰Belgium consists of 581 municipalities and 19,781 statistical sectors (Vademecum Statistische sectoren, 2012). Statistical sectors have a surface of 1.5 km² and 539 inhabitants, on average.

Table 1: Enrollment in elite and non-elite schools

		Elite school	Non-elite school
<i>Total enrollment</i>		26.7%	73.3%
<i>Characteristics of students</i>			
Gender	Male	48.6%	49.9%
	Female	51.4%	50.1%
Repeated during primary school	Yes	6.0%	13.1%
	No	94.0%	86.9%
Degree of the mother	Higher education degree	55.8%	32.4%
	High school degree	32.8%	43.1%
	No high school degree	11.4%	24.5%
Language at home	Dutch	94.3%	95.1%
	Other language	5.7%	4.9%
Household income	High income ^a	85.9%	78.5%
	Low income	14.1%	21.5%
<i>Neighborhood characteristics of students (average)</i>			
Median household income (in 1000 euro)	Municipality	19.9	19.7
	Statistical sector	20.8	20.1
Belgian nationality	Municipality	95.6%	95.8%
	Statistical sector	96.0%	96.1%
Higher education degree	Municipality	28.0%	25.4%
	Statistical sector	29.7%	25.1%
High school degree	Municipality	29.6%	30.2%
	Statistical sector	29.6%	30.5%
No high school degree	Municipality	42.4%	44.4%
	Statistical sector	40.7%	44.4%
Population density (in 100 inhabitants per km ²)	Municipality	8.2	6.6
	Statistical sector	26.7	23.4
<i>Distance to school (km, average)</i>			
Distance to closest elite school		3.3	5.4
Distance to closest non-elite school		2.4	2.3
Observations		29,762	81,809

Note: 111,571 students starting secondary education in 2003 or 2004. Percentages are conditional on school choice. Median household income is measured in 2004 and other neighborhood characteristics are measured in 2001.

^aNot eligible for study grant.

characteristics at the level of the statistical sector thus reflect the residential neighborhood, while characteristics at the level of the municipality capture amenities in a broader region.

Table 1 shows that students who attend elite schools live in neighborhoods with a higher median income and more highly educated individuals. Students in elite schools live in neighborhoods with a similar share of Belgians as students in non-elite schools. Finally, average population density is higher for students attending elite schools because relatively more elite schools are located in cities. We also observe the location of students and schools.¹¹ Average distances to both types of schools are small (2 to 5 km) and the average distance to elite schools is smaller for the group of students choosing an elite school (3 instead of 5 km).¹²

Table 2 shows the track choice in the second year and the outcome variables that we consider in the analysis. 93.6% of students who started at an elite school enrolled in the

Table 2: Track choice and study outcomes by initial school choice

	Elite school	Non-elite school
Academic track in grade 8	93.6%	55.3%
High school diploma	94.3%	90.8%
High school diploma without study delay	77.5%	70.0%
Downgrading ^a	27.1%	35.5%
Observations	29,762	81,809

Note: 111,571 students starting secondary education in 2003 or 2004. Percentages are conditional on school choice.

^aDummy variable equal to 1 when a student switches from a high to a low track during secondary education.

academic track in the second year of secondary education. However, only 55.3% of students who started at a non-elite school enrolled in the academic track. Note that students in elite schools have to leave the school if they want to start in a non-academic track.

Our main outcome variable is graduating from high school. We define this variable by looking at the data for these students until 2011 (cohort of 2003) and 2012 (cohort of 2004). Since it takes a minimum of six years to graduate from high school, we allow for a study

¹¹We observe the location of the students at the level of the statistical sector. We compute the distance from the center of the statistical sector to the exact location of the school.

¹²Both types of schools are within commuting distance for almost all students. Distance to a non-elite school is 2.4 km, on average, with a standard deviation of only 1.9 km. Distances to elite schools are greater: 4.8 km, on average, with a standard deviation of 3.8 km. Nevertheless, at the 95th percentile, there is still a reasonable distance to travel of 11.6 km.

delay of up to three years.¹³ Students who started at elite schools perform better: 94.3% of them graduate compared to 90.8% of the students who started at a non-elite school. Study delay is common. Only 77.5% of students starting in an elite school and 70.0% of students starting in a non-elite school graduate on time. Finally, downgrading (i.e., a switch from a higher to a lower track) is more common in non-elite schools: 35.5% of the students starting at non-elite schools downgraded at least once, compared to only 27.1% in elite schools. As for the initial track choice, downgrading requires switching to a different school for students in elite schools.

3 Empirical framework

We study the causal effect of starting at an elite school on high school graduation. To address the potential self-selection of high ability students in elite schools, we use distance as an instrument for school choice. We first specify a potential outcome framework. We then discuss how to build a semi-parametric estimator for the average treatment effects, using the Marginal Treatment Effect (MTE) framework of Heckman and Vytlacil (2005).¹⁴ Next, we show how parametric and distributional assumptions can overcome common support problems of the semi-parametric approach and how the factor structure of Aakvik et al. (2005) provides additional insights into the distribution of treatment effects.

3.1 Potential outcomes

Let Y_{1i} be the outcome of student i in the case of treatment, starting at an elite school ($D_i = 1$), and Y_{0i} the outcome of student i if he or she starts at a non-elite school ($D_i = 0$). For each student i , we observe only the realized outcome Y_i in the observed state:

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i} \tag{1}$$

with $Y_{1i} = \mu_1(X_i, U_{1i})$ the outcome in case of treatment, and $Y_{0i} = \mu_0(X_i, U_{0i})$ in the case of no treatment. X_i is a vector of observed characteristics of students and U_{0i} and U_{1i} are unobserved random variables.

Let selection into treatment be given by

$$D_i = 1(V_i \leq \mu_D(X_i, Z_i)) \tag{2}$$

¹³0.3% of the students are still enrolled in secondary education in the last year of our dataset and have not yet obtained their degree at the end of this year. We consider them as drop-outs in the main analysis. We obtain similar results when omitting these students or assigning them to the group of high school graduates (see Supporting Information Appendix B).

¹⁴See also Cornelissen et al. (2016) for an overview on how to apply this method.

with $1(\cdot)$ an indicator function = 1 if the expression between brackets is true, $\mu_D(X_i, Z_i)$ an arbitrary function of observed characteristics X_i and an instrument Z_i . V_i is the unobserved cost of treatment for student i with F_V its cumulative distribution function. We can then apply a probability normalization such that $D_i = 1(U_{Di} \leq P(X_i, Z_i))$ with $P(X_i, Z_i) \equiv F_V(\mu_D(X_i, Z_i))$ the propensity score and $U_{Di} \equiv F_V(V_i)$ an unobserved cost of treatment that is uniformly distributed on the unit interval, regardless of the distribution of V .¹⁵ Our main interest is in the Average Treatment Effect: $ATE = E[Y_{1i} - Y_{0i}]$, the Average Treatment effect on the Treated: $ATT = E[Y_{1i} - Y_{0i} | D_i = 1]$ and the Average Treatment effect on the Non-Treated: $ATNT = E[Y_{1i} - Y_{0i} | D_i = 0]$.

3.2 Semi-parametric estimation of average treatment effects

To allow for heterogeneous effects, we follow the framework of Heckman and Vytlačil (2005) by using MTEs as building blocks for constructing the ATT, ATE and ATNT. The MTE is the effect of attending an elite school for students with observable characteristics $X_i = x$ and unobserved cost of treatment $U_{Di} = u_D$:

$$MTE(x, u_D) = E[Y_{1i} - Y_{0i} | X_i = x, U_{Di} = u_D].$$

We can interpret the treatment effect at low values of u_D as the effect for students who have a low unobserved cost; that is, a high unobserved preference for attending an elite school. The effect at high values is the effect for students with high unobserved costs, or low unobserved preferences. A weighted average of the MTEs then yields estimates of the ATE, ATT and ATNT (Heckman and Vytlačil, 2005).

As explained in Cornelissen et al. (2016), the minimal set of assumptions we need for the estimation of the MTEs is similar to the set of assumptions needed for the LATE interpretation of the 2SLS estimator (Imbens and Angrist, 1994). In section 4 we discuss the validity of these assumptions in the context of our application based on empirical evidence. In the rest of the paper, we maintain the following assumptions on the instrument Z_i :

Condition 1 (*Relevance*) Z_i is a random variable such that the propensity score $P(X_i, Z_i)$ is a nontrivial function of Z_i .

Condition 2 (*Independence*) (U_{0i}, U_{1i}, U_{Di}) are independent of Z_i conditional on X_i .

The first condition assures that the instrument is relevant in the school choice. The second condition assures that the instrument has no direct impact on study outcomes. Vytlačil

¹⁵The distribution has to be absolutely continuous with respect to the Lebesgue measure.

(2002) also notes that the selection model in (2) is equivalent to the monotonicity condition of the LATE literature. In our context, this means that students should perceive the distance to school as a cost and not as a benefit.

Estimation of the MTEs can then proceed using local IV by determining how Y_i changes with $P(X_i = x, Z_i = z)$ in small neighborhoods around each value of $P(x, z)$ (Heckman and Vytlacil, 1999). To identify the entire $MTE(x, u_D)$ function, we need common support, i.e. we need sufficient observations of both treated and non-treated students at each value of $P(x, z)$. Therefore, we follow a semi-parametric approach where the control variables enter the outcome equation linearly, but the functional form of the MTE curve (across control variables) is not restricted (see Section 4.3 in Cornelissen et al. (2016)).

In Section 4, we show that we can identify $MTE(x, u_D)$ for all $u_D \leq 0.81$.¹⁶ This allows us to interpret how the costs of treatment relate to heterogeneity in treatment outcomes. This is sufficient to estimate the ATT because the weights in the calculation of the ATT are 0 for $u_D > 0.81$.¹⁷ Nevertheless, this approach does not allow estimation of the ATNT and ATE with our data. Therefore, we discuss a model that introduces functional form and distributional assumptions.

3.3 Parametric estimation of average treatment effects

We follow Manski et al. (1992) and impose parametric and distributional assumptions to model treatment and potential outcomes. The structure is sufficiently flexible to allow for heterogeneous treatment effects and does not impose the structure of a Roy model, that is, students do not need to select on gains of treatment. We use the following functional form assumptions:

$$\begin{aligned} D_i &= 1(\gamma Z_i + \delta X_i - V_i \geq 0), \\ Y_{0i} &= 1(\beta_0 X_i - U_{0i} \geq 0), \\ Y_{1i} &= 1(\beta_1 X_i - U_{1i} \geq 0), \end{aligned} \tag{3}$$

with γ , δ , β_0 and β_1 parameters to estimate. We assume that the error terms (V_i, U_{0i}, U_{1i})

¹⁶We estimate the model in STATA using the user-written command 'mtefe' (Andresen, 2018).

¹⁷Equation (28) in Cornelissen et al. (2016) shows that we can calculate the ATT as follows: $ATT = \frac{1}{N} \sum_{i=1}^N \frac{p_i}{\bar{p}} X_i (\beta_1 - \beta_0) + \sum_{u=1}^{100} \frac{\Pr(p_i > u/100)}{100\bar{p}} E(U_{i1} - U_{i0} | U_{iD} = u/100)$ with p_i the predicted propensity score for i and $E(U_{i1} - U_{i0} | U_{iD} = u/100)$ the unobserved component of the MTE. We find that $\Pr(p_i > u/100)$ is approximately 0 for all $u > 81$.

are jointly normal with a mean-zero vector and correlation matrix Ω :

$$\Omega = \begin{pmatrix} 1 & \rho_0 & \rho_1 \\ & 1 & \rho_{10} \\ & & 1 \end{pmatrix}. \quad (4)$$

Because Y_{0i} and Y_{1i} are never observed simultaneously, the joint distribution of (U_{0i}, U_{1i}) and thus their correlation ρ_{10} is not identified. We can however estimate correlations between V_i and U_{0i} : ρ_0 and between V_i and U_{1i} : ρ_1 .¹⁸ Since the joint distribution of each outcome equation and the selection equation is identified, we have sufficient information to calculate the ATE, ATT and ATNT for each value of x (Lokshin and Sajaia, 2011). We then integrate over its empirical distribution to construct the overall averages. We also use this model to identify average marginal effects by calculating how the treatment effects differ when one of the covariates changes value, keeping the other covariates constant.

3.4 Factor model and distributional treatment effects

While ρ_{10} was not needed for average effects, it is required to learn about the distribution of treatment effects. In particular, we are interested in the proportion of students experiencing positive or negative treatment effects. Aakvik et al. (2005) show that identification can be achieved by imposing a factor structure on the error terms.¹⁹ They follow a similar structure as in (3), but impose the following factor structure on the error terms:

$$\begin{aligned} V_i &= -\theta_i + \varepsilon_{Di}, \\ U_{0i} &= -\alpha_0\theta_i + \varepsilon_{0i}, \\ U_{1i} &= -\alpha_1\theta_i + \varepsilon_{1i}, \end{aligned} \quad (5)$$

with $\theta, \varepsilon_D, \varepsilon_0$ and ε_1 all independently distributed (e.g. $N(0, 1)$); that is, the correlation between error terms enters exclusively through one common factor, θ_i . This restricts the correlation between unobservables of the potential outcomes to be driven by the same reason for choosing the school. θ_i can be interpreted as unobserved ability or social status. If the strongest students select into treatment, we should see $\alpha_0 > 0$ and $\alpha_1 > 0$. With selection

¹⁸We estimate the model in STATA using the user-written command 'switch_probit' (Lokshin and Sajaia, 2011).

¹⁹The non-identification of ρ_{10} was also addressed by Poirier and Tobias (2003) without imposing a factor model. They make use of an algorithm in Koop and Poirier (1997), who use the restrictions that come from the positive definiteness of the covariance matrix of the error terms to learn about the value of ρ_{10} . The authors use a Bayesian procedure to estimate its distribution but warn that the results can depend on the prior that is chosen for ρ_{10} .

on gains, we would have $\alpha_1 > \alpha_0$. However, if unobserved preferences for treatment are correlated with behavior that reduces the outcome, we can have $\alpha_1 < \alpha_0$, with α_1 possibly negative. An alternative interpretation of θ_i is that it captures the local quality of schools. If elite schools in some neighborhoods are relatively better, we would expect $\alpha_1 > \alpha_0$ if students that live in these areas are more likely to attend them. With normally distributed $\theta, \varepsilon_D, \varepsilon_0$ and ε_1 , the model is identical to the one we proposed in the previous subsection with $\rho_j = \frac{\alpha_j}{\sqrt{2}\sqrt{1+\alpha_j^2}}$ for $j \in \{0, 1\}$.²⁰ The previously unidentified correlation, ρ_{10} , is now identified, as it only depends on the two factor loadings: $\rho_{10} = \frac{\alpha_0\alpha_1}{\sqrt{1+\alpha_0^2}\sqrt{1+\alpha_1^2}} = 2\rho_0\rho_1$. This identity can then be used to calculate distributional treatment effects.²¹ Note that this factor structure was only needed to identify ρ_{10} and therefore has no impact on the estimates of the ATT, ATNT, ATE, or marginal effects.

To estimate this model, we proceed as follows: we first estimate (3) with error structure (4) using maximum likelihood to calculate the ATT, ATNT, ATE, and the average marginal effects of observed characteristics. We then apply the additional structure on the error terms (5) to calculate distributional treatment effects.

4 Instrument validity

We use the relative distance to an elite school as an instrument for school choice. We define the relative distance as the distance to the closest non-elite high school, subtracted by the distance to the closest elite high school (in kilometers). The use of geographical variation as an instrument for school choice has been proposed by Card (1995) and has also been used more recently (see, for example, Barrow et al. (2015), Carneiro et al. (2016) and Nybom (2017)). In this section, we discuss the conditions for the instrument.

4.1 Relevance

The first condition implies that distance should have a strong impact on school choice. Table 3 shows the results of the first stage of a 2SLS regression.

The first specification controls for the neighborhood characteristics summarized in Table 1, fixed effects for each of the five provinces and for each level of urbanization of the

²⁰Note that normalizations are different. Instead of normalizing the variances of V_i, U_{0i} , and U_{1i} to be 1, the model of Aakvik *et. al* (2005) implies variances of respectively 2, $\alpha_0^2 + 1$ and $\alpha_1^2 + 1$; that is, our estimates should be multiplied by the square roots of these numbers to translate them to their model.

²¹With the three correlation parameters identified, we can take draws from the estimated error distribution, calculate the potential outcomes and simulate treatment effects. We also directly estimate the factor model using the 'gsem' command in STATA and obtain identical results for the treatment effects.

Table 3: First stage: choosing an elite school

Variables	(1)		(2)	
	Coef.	St. error	Coef.	St. error
Relative distance	0.024*	(0.001)	0.024*	(0.001)
Mother no high school degree ^a			-0.181*	(0.004)
Mother high school degree ^a			-0.123*	(0.003)
No Dutch at home			0.063*	(0.007)
Low income			-0.029*	(0.003)
Male			-0.011*	(0.003)
Repeated			-0.121*	(0.004)
Neighborhood characteristics		yes		yes
Observations		111,571		111,571
R-squared		0.131		0.169
F-stat excl. instr.		867		878

Note: Standard errors are corrected for clustering within the statistical sector. Neighborhood characteristics include population density, median household income, the share of Belgians and average educational level at the municipality and the statistical sector. Both regressions additionally control for province fixed effects and indicators for the level of urbanization of the municipality.

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

municipality. The second specification additionally controls for student characteristics. The relative distance has the expected positive sign and is highly significant, resulting in a high F-statistic of the exclusion restriction. Students (or their parents) are sensitive to distance to school as an additional kilometer to a non-elite high school (or an equal reduction in the distance to the elite school) makes them 2.4 percentage points more likely to choose the elite school. This estimate does not change when controlling for observed student characteristics (which is what we would expect from a valid instrument). The control variables indicate that selection into elite schools is not random, as disadvantaged students are less likely to choose an elite school. Students who do not speak Dutch at home are the only exception. Despite their lower expected outcomes, they are more likely to choose an elite school.²²

²²A possible explanation is that this variable groups students from diverse migrant backgrounds, both low and high skilled. While low skilled migration was very common in Flanders for industrial production during the 20th century, high skilled migration is also important because of Flanders' proximity to Brussels, the capital of the European Union.

4.2 Independence

The second condition implies that differences in distances must be independent of unobservables in the outcome equation. While distances can be problematic in several institutional contexts, we argue that this is not the case in Flanders.²³ Unrestricted school choice is protected by the Belgian constitution (art. 24, §1) and most students live close to several schools. They can choose between 667 schools of which 131 are elite schools. For the median student, six schools are located within a 5 km radius.²⁴ Students can also benefit from bike lanes or inexpensive public transportation. Parents are still expected to take neighborhood characteristics into account when deciding where to live. These characteristics might be correlated with the presence of an elite school. Controlling for neighborhood characteristics addresses this concern.

Although we cannot show that travel distance is independent of unobserved student characteristics, it would be reassuring if it is not correlated with observed student characteristics. In Table 4, we show the results of an OLS regression of relative distance on student characteristics. We conclude that after controlling for neighborhood characteristics, we do not find evidence of selection into neighborhoods with elite schools based on observed student characteristics. A joint significance test is strongly rejected, and each covariate is precisely estimated around 0.²⁵

²³Altonji et al. (2005) show that distance is not a valid instrument to identify the effect of attending a Catholic high school in the US.

²⁴In Supporting Information Appendix B, we show that our main results are similar when we restrict the sample to students who have at least four schooling options in a 5 km radius and who have an elite school among the four closest options.

²⁵Supporting Information Appendix C shows that it is particularly important to control for intraregional differences, captured by dummy variables for the provinces and the levels of urbanization. When we do not control for neighborhood characteristics, the effect of student characteristics on travel distance is larger and significant. This is because we find that academic outcomes are lower in urban areas. At the same time, distances to different types of schools are smaller in urban areas, which implies that elite schools are more accessible for students living in cities. Not controlling for neighborhood characteristics then leads to a downward bias in the estimates of the treatment effects.

Table 4: Regression of relative distance on student characteristics

Variables	Coef.	St. error
Mother no high school degree ^a	-0.055	(0.033)
Mother high school degree ^a	-0.023	(0.024)
No Dutch at home	0.040	(0.062)
Low income	-0.046	(0.030)
Male	-0.010	(0.019)
Repeated	-0.025	(0.032)
Neighborhood characteristics		yes
F-test individual characteristics	1.28 (p=0.263)	
Observations	111,571	

Note: Standard errors are corrected for clustering within the statistical sector. Neighborhood characteristics include population density, median household income, the share of Belgians and average educational level at the municipality and the statistical sector. The regression additionally controls for province fixed effects and indicators for the level of urbanization of the municipality.

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

4.3 Common support

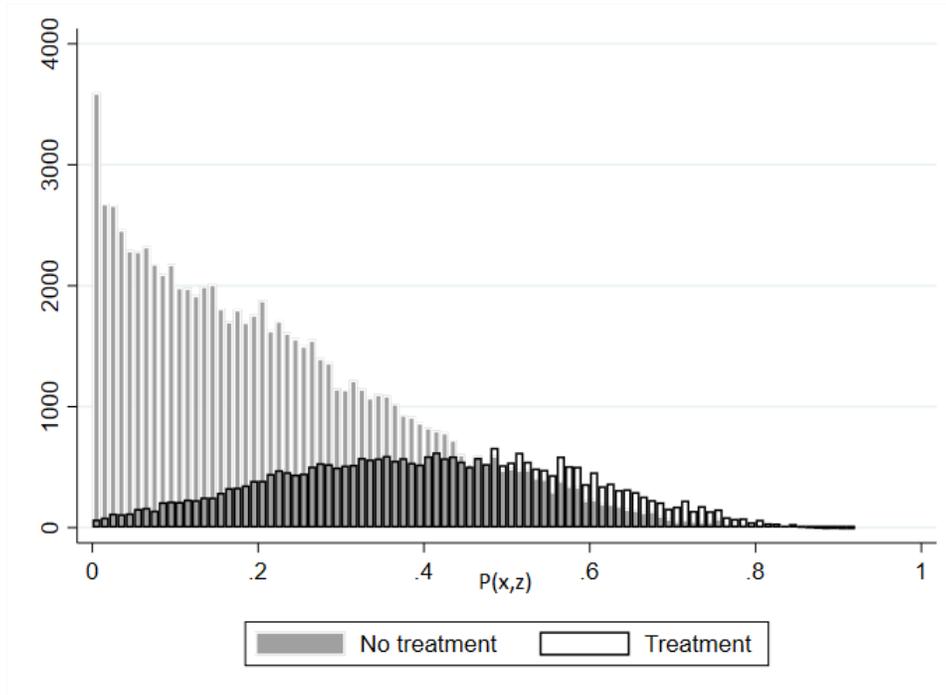
This last condition implies that for each value of the propensity score $P(x, z)$, there should be treated and non-treated individuals. This condition is needed only for the semi-parametric approach. We estimate the propensity score using a probit regression and plot the histogram of common support in Figure 1.

The common support assumption is satisfied for values of $P(x, z)$ up until 0.81. For higher values of $P(x, z)$, there are almost no treated or non-treated students and the MTE cannot be identified.²⁶ This allows us to compute the ATT because the weights in the calculation of the ATT become 0 for $u_D > 0.81$.²⁷ Therefore, we do not need estimates of the MTE for $u_D > 0.81$ and thus we do not need common support at $P(x, z) > 0.81$.

²⁶Each percentile up until 0.81 contains at least 10 treated and 10 non-treated students. For most percentiles above 0.81, this condition is not satisfied so we do not report an MTE here.

²⁷In practice, the weights are non-zero but very small (see Figure D1).

Figure 1: Common support



Note: This figure plots the number of treated and non-treated students for each percentile of the propensity score (111,571 students).

5 Empirical results

We first discuss the estimation results and then proceed to the implied treatment effects.

5.1 Estimation results

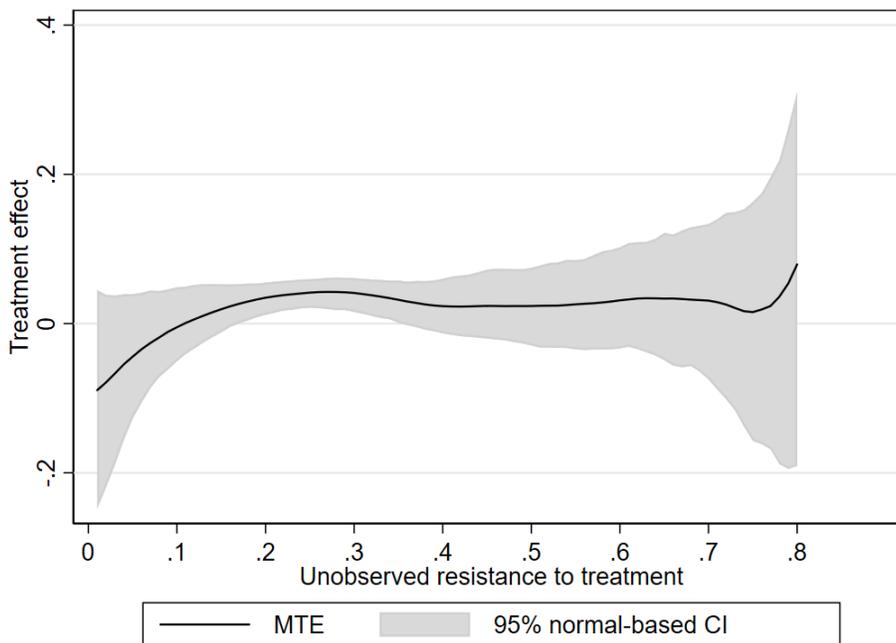
We estimate both a semi-parametric and parametric model to study the impact of elite schools on high school completion.²⁸ For the semi-parametric model, we show the estimation results of the observable part of the treatment effect in Table D2 in Supporting Information Appendix D. In Figure 2, we plot the unobservable part; that is, the MTE curve.²⁹ On the horizontal axis, we present u_D . Students with a low u_D have a low unobserved cost (or a high unobserved preference) for elite schools and are therefore more likely to be treated. Figure

²⁸OLS finds a positive effect from treatment, but this effect decreases when controlling for background characteristics. When we control for self-selection with a 2SLS regression, the treatment effect turns out to be statistically insignificant (see Supporting Information Appendix D, Table D1).

²⁹The bandwidths of the local polynomial regressions are determined by the asymptotically optimal constant bandwidth (Fan and Gijbels, 1996).

2 shows that the treatment effect remains constant and rarely significantly differs from zero. This does not mean that treatment effects are constant. Many of the control variables are capturing heterogeneity in returns. As expected from the MTE curve, we cannot reject the hypothesis of no unobserved heterogeneity in the treatment effect (p-value = 0.424) but we reject that treatment effects are homogeneous for observed characteristics (p-value = 0.000).

Figure 2: Obtaining a high school degree: MTE



Note: MTEs are calculated based on 111,571 students using a local polynomial regression of degree 2 with an Epanechnikov kernel. Standard errors are computed with a bootstrap procedure (250 replications) and clustered within the statistical sector. P-value no observable heterogeneity (0.000), p-value no unobservable heterogeneity (0.424).

Table D3 in Supporting Information Appendix D contains the estimates of the parametric model. Here we find a statistically significant impact of unobservables. The unobservable in the non-treated outcome is positively correlated with the unobservable in the selection equation ($\rho_0 = 0.193$). This is consistent with the idea that stronger students select into elite schools. However, for the treated outcome we find a negative correlation ($\rho_1 = -0.185$).³⁰ This suggests that these stronger students lose their advantage by attending an elite school.

³⁰As explained in section 3.4, we can also use these correlations to identify the factor loadings in the model of Aakvik et al. (2005): $\alpha_0 = 0.283$ and $\alpha_1 = -0.271$.

Note that the unobserved factor captures only what observable characteristics fail to explain in the correlation between preferences for elite schools and high school graduation. A potential explanation for the different signs of ρ_0 and ρ_1 is that an important number of students with a high preference for elite schools is at risk of not being able to pursue the academic track. They could have above average ability, leading to better outcomes in non-elite schools. At the same time, they are not top performers implying that they have a risk of failing in the academic track. This can have detrimental effects on the probability to graduate from high school if their high preference makes them unwilling to switch to another school. In section 6 we investigate the interaction between school and track choice and find further evidence for this explanation.³¹

5.2 Average and distributional treatment effects

ATT, ATNT and ATE

Table 5 compares the estimates of the average treatment effects of the semi-parametric and parametric models. From the semi-parametric model, we compute a statistically insignificant ATT of 0.4 percentage points.

Table 5: Obtaining a high school degree: average treatment effects

	Semi-parametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error
ATT	0.004	(0.019)	-0.016	(0.008)
ATNT	<i>Not identified^a</i>		0.051*	(0.010)
ATE	<i>Not identified^a</i>		0.033*	(0.008)

Note: Average treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications.

* Statistical significance at 5% level.

^a Not identified due to insufficient common support.

Despite the extra structure of the parametric model, we still obtain a statistically insignificant ATT (−1.6 percentage points). We derive a significantly positive ATE of 3.3 percentage points and an ATNT of 5.1 percentage points. These results imply that students who do not attend elite schools would, on average, be 5.1 percentage points more likely to

³¹Reverse selection on unobserved gains is not uncommon in the literature as it has also been established in Aakvik et al. (2005) on the effects of a rehabilitation program on employment and Cornelissen et al. (2018) on the effects of universal child care on school outcomes.

graduate from high school when they would have chosen an elite school. In contrast, students who attend elite schools do not, on average, experience an effect from treatment.³²

Average marginal effects of student characteristics

The average treatment effects can still mask a lot of heterogeneity within both groups of treated and non-treated students. Therefore, we consider the effect of observable characteristics and calculate distributional treatment effects. We report the average marginal effects of each background variable in Table 6, as computed from the parametric model. They report the change in the ATE when the value of the control variable changes from 0 to 1 for all individuals in the sample. As these effects differ between individuals, we evaluate them for each student in the sample at their actual realizations of all other variables and report the mean effect.

The average marginal effects are suggestive of reverse selection on gains. Table 6 shows that choosing an elite school has larger positive effects for students whose mother did not complete high school and students who repeated at least one grade in primary school. However, these students are less likely to choose an elite school (see Table D3 in Supporting Information Appendix D). As explained before, the estimates also reveal such a pattern for unobservables. An example that is consistent with this story is that of a low ability student from a high social class who is strongly encouraged by his or her parents to attend an elite school but would be more likely to graduate from high school when choosing a non-elite school. The reverse is also possible. High ability students with a low social status might not choose an elite school, even if they would be more successful in this type of school. As we find the strongest marginal effects for the educational level of the mother, we believe this is a plausible story.

Distributional treatment effects

Distributional treatment effects identify the fraction of students who experience or would have experienced (in case of no treatment) a positive or negative treatment effect. The second panel of Table 6 shows the distributional treatment effects for the total population of students, the group of students who started at elite schools and the group of students who

³²We repeated the analysis with an alternative outcome variable: obtaining a high school degree without study delay (i.e., within six years of studying). Other studies have found that grade retention has a negative effect on performance in the final years of high school (Cockx et al., 2018) and on the student's beginning-of-career wage (Brodaty et al., 2013). Similar to our main outcome, we obtain a statistically insignificant ATT but we find even stronger positive effects for the ATNT. The results can be found in Supporting Information Appendix D Tables D4, D5 and D6, and the MTE in Figure D2.

Table 6: Obtaining a high school degree: average marginal effects and distributional effects

	Coef.	St. error
<i>Average marginal effects of student background</i>		
Mother has no high school degree ^a	0.049*	(0.010)
Mother has high school degree ^a	0.014*	(0.005)
No Dutch at home	-0.011	(0.008)
Low income	-0.005	(0.004)
Male	0.010	(0.006)
Repeated	0.028*	(0.012)
<i>Distributional treatment effects^b</i>		
<i>Among all students</i>		
% benefit from elite school	7.232*	(0.239)
% suffer from elite school	3.941*	(0.616)
<i>Among students in elite school</i>		
% benefit from elite school	3.680*	(0.702)
% suffer from elite school	5.220*	(0.176)
<i>Among student in non-elite school</i>		
% benefit from elite school	8.549*	(0.233)
% suffer from elite school	3.468*	(0.837)

Note: Average marginal effects and distributional treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. Average marginal effects of control variables report the change in the ATE when the value of the control variable would change from 0 to 1. Distributional effects are calculated by simulating the error terms of the potential outcome framework (100 draws for each student).

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

^b Requires identification of ρ_{10} using factor structure Aakvik et al. (2005).

started at non-elite schools. In total, 7.2% of students experience a positive treatment effect. Note that treatment effects are large, as they involve a switch from not obtaining a degree to obtaining a degree (or vice versa). A smaller fraction of 3.9% would graduate from high school only if they started at a non-elite school. Among the group of treated students, we compute that 5.2% of students experience a negative treatment effect and that only 3.7% experience gains from attending an elite school. This is consistent with our previous finding

that students in non-elite schools would experience the most positive treatment effects. While 8.5% of students in non-elite schools would experience a positive treatment effect, only 3.5% would lose from attending an elite school. Therefore, we can conclude that 7.6% of students would have been better off in the other school type.³³

The finding that students can be negatively affected by an elite school is not new. Pop-Eleches and Urquiola (2013) find that parents reduce their effort when their child enters an elite school. They also find that relatively weak students in better-ranked schools feel more marginalized and insecure than similar students in other schools. A negative effect can also be explained by differences in grading standards. Schools have some autonomy in deciding grade progression and excluding students from certain programs. Therefore, it is possible that elite schools more often require weaker students to repeat a grade to achieve the required schooling level, which could demotivate them and result in drop out.

Another potential explanation for the heterogeneity in the effects is that elite schools differ from non-elite schools in their supply of tracks. If students are not aware of their ability from the start of high school, a wrong track choice may occur. This could be problematic when they are in an elite school because of the higher costs of switching to a lower track. We will explore this channel further in Section 6.

6 Explaining heterogeneity in the treatment effect by track choices

In this section, we explain our two main findings. First, attending an elite school, on average, increases degree completion, but only for the non-treated students. This implies that there is a reverse selection on gains. Second, there is substantial heterogeneity in the treatment effect within the group of treated and non-treated students. Consequently, several students in both elite and non-elite schools experience a positive or negative treatment effect from attending an elite school.

In the first part of this section, we explain these results by extending the factor model of Section 3.4 with the track choice in the second year of high school (grade 8).³⁴ We use this model to isolate the effect of the tracking decision by performing two sets of simulations. First, we simulate the track choice that students would make in each type of school. Second,

³³This is the result of a weighted sum: 5.2% of the 26.7% in elite schools suffer, while 8.5% of the 73.3% in non-elite schools would gain.

³⁴The “official” track distinction starts a year later (grade 9) but study programs in grade 8 are already aligned with the track. We therefore also show the results using the track in grade 9 instead and find similar effects (see Supporting Information Appendix D, Table D9 and Table D10).

we simulate the outcomes in each school when students are forced to enroll in a specific track.

In the second part of this section, we study how elite schools affect the probability of switching to a lower track at a later age (downgrading).

6.1 Endogenous tracking decisions

Let $A_{ji} = 1(\sigma_j Z_i^{track} + \kappa_j X_i - U_{ji}^{track} \geq 0)$ be an indicator equal to one when a student enrolled in the academic track, with $j = 0$ when student i entered a non-elite school and $j = 1$ if student i entered an elite school. Track choices depend on the same set of covariates X_i that affects the choice of the type of school, but also the relative distance to the nearest school that offers the academic track Z_i^{track} .³⁵ We further extend the model to have four instead of two potential outcomes (conditional on each school type and track).³⁶ The error terms are still generated by a factor model.

The estimates can be found in Supporting Information Appendix D, Tables D7 and D8. Table 7 shows the estimated treatment effects of the model. First, we show that the treatment effects of choosing an elite school on degree completion in secondary education are similar as in the main analysis in Table 5. Next, we find that choosing the academic track raises the probability of graduating from high school for the treated (+6.8 percentage points) and the non-treated (+4.5 percentage points). Finally, choosing an elite school makes students, on average, 23.1 percentage points more likely to enroll in the academic track. This average effect is largely driven by non-treated students whose impact is more than three times larger than the impact on treated students. This can be explained by the correlation between preferences for elite schools and preferences for the academic track.³⁷ While many treated

³⁵We define Z_i^{track} as the distance to the closest school that offers a non-academic track, subtracted by the distance to the closest school that offers an academic track.

³⁶We also add Z_i^{track} as a covariate in the school choice equation to allow students to consider the distance to the academic track when making their schooling decision. Note that the large number of neighborhood characteristics requires us to constrain some of their effects. We assume that the parameters of neighborhood characteristics in the track choice equations do not depend on the initial school choice. As in the baseline model, we allow their effects on the outcome to differ between both school choices, but we restrict them to be the same over track choices. Note that we still allow individual characteristics (observed and unobserved) to enter in a fully flexible way.

³⁷Table D7 shows that observable characteristics predicting students to enroll in an elite school are the same as those predicting them to enroll in the academic track. This pattern also holds for unobservables. We find a factor loading of 0.870 for students in non-elite schools and 0.251 for students in elite schools. As the factor loading of attending an elite school is normalized to 1, a positive load on the track choice points to a positive correlation between unobserved preferences for elite schools and unobserved preferences for the academic track.

students would enroll in the academic track in any case, this is not true for the non-treated students who are more encouraged to enroll in the academic track if they attended an elite school.

Table 7: Extended model: average treatment effects

	Coef.	St. error	Coef.	St. error	Coef.	St. error
Treatment	Elite school		Academic track		Elite school	
Outcome	High school degree		High school degree		Academic track	
ATT	-0.008	(0.007)	0.068*	(0.014)	0.076*	(0.008)
ATNT	0.049*	(0.011)	0.045*	(0.013)	0.287*	(0.036)
ATE	0.033*	(0.008)	0.061*	(0.012)	0.231*	(0.028)

Note: Average treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed by sampling 250 draws from the estimated distribution. For each draw of the estimates, we calculate the treatment effects by simulating the error terms of the potential outcome framework (100 draws for each student).

* Statistical significance at 5% level.

We allow for two channels through which elite schools can affect degree completion. First, elite schools can have a direct effect on degree completion which could be caused by better peers or teachers, for example. Second, elite schools encourage students to start in the academic track and thereby make them more likely to obtain a degree. In Table 8, we isolate the direct effect of choosing an elite school by simulating degree completion for treated and non-treated students, while keeping the track choice fixed.

Table 8: Extended model: average treatment effects conditional upon track choice

	No restrictions		Only academic track		No academic track	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
ATT	-0.008	(0.007)	-0.003	(0.006)	-0.130*	(0.036)
ATNT	0.049*	(0.011)	0.042*	(0.012)	0.000	(0.025)
ATE	0.033*	(0.008)	0.030*	(0.009)	-0.035	(0.024)

Note: Average treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed by sampling 250 draws from the estimated distribution. For each draw of the estimates, we calculate the treatment effects by simulating the error terms of the potential outcome framework (100 draws for each student).

* Statistical significance at 5% level.

The results in the first column are again the estimated treatment effects of choosing an elite school on high school graduation when the track choice is not restricted. In the second column, we simulate how treatment effects would change if every student would enroll in the academic track. While the ATT is identical to the baseline case, the ATNT is slightly smaller.³⁸ In this simulation, elite schools no longer incentivize students to enroll in the academic track. The positive treatment effect for the non-treated suggests that elite schools not only raise the probability to choose the academic track but also success.

In the third column, we simulate how treatment effects would change when the academic track cannot be chosen. Note that in this case, students enrolled in elite schools have to move to another school after the first year. While the non-treated experience no effect, we find a large and negative ATT of -13.0 percentage points. This result shows that students with a high preference for elite schools suffer when they are not able to start at the academic track and have to switch to a different school. For non-treated students, this is less of an issue. They would also need to switch schools but rather it would be a switch to their preferred school in any case so it is not considered to be as costly.

We conclude that differences in tracking decisions can explain the reverse selection on gains. First, choosing an elite school raises the likelihood of enrolling in the academic track. Second, elite schools make students more likely to persist and succeed in this track. Third, this is the case only for non-treated students as treated students tend to enroll in the academic track, independent of the type of school they choose in the first year. Finally, we find that treated students are negatively affected by an elite school if they are not able to pursue the academic track. Their preference for elite schools makes them unwilling to switch to another school to downgrade to a lower track, which could help them avoid study delay and dropout.

6.2 Downgrading

The model in the previous subsection shows the impact of elite schools on the initial track decision, as well as how treated and non-treated students react differently to the non-availability of a track. In this section, we show that this initial response also persists in the higher grades, by repeating our main analysis for an alternative outcome variable: downgrading.³⁹

From the estimated treatment effects in Table 9, we can conclude that students with a high preference for elite schools are more prepared to stay in the academic track if they attend an elite school. Students with lower preferences experience no effect because they

³⁸We also find a smaller ATNT when repeating our main analysis on the sample of students who choose the academic track in the second grade of high school (see Supporting Information Appendix B).

³⁹The estimates can be found in the Supporting Information Appendix D, Tables D11 and D12 and the MTE in Figure D3.

are more willing to change schools to switch tracks. This can also be seen in the average marginal effects (see Table 10). For most variables, such as low education of the mother and having repeated a grade in elementary school, we see that students benefitting more from elite schools (see Table 6) are also the ones who have positive marginal effects on the downgrading decisions. This implies that the negative ATE on downgrading of -4.1 percentage points is mostly coming from students who do not improve their chances of obtaining a high school degree by attending an elite school. However, students who do experience improvements do not let their downgrading decision depend on the type of school. In an elite school, they make use of the opportunity to be more successful after choosing the academic track (see section 6.1), but if they need to switch at a later date, they are still willing to do so. Note that this high preference does not necessarily need to apply to the student. They can also be strongly encouraged by their parents to attend and/or stay in an elite school, even when teachers give the opposite advice.

Table 9: Downgrading: average treatment effects

	Semi-parametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error
ATT	-0.133*	(0.021)	-0.109*	(0.020)
ATNT	<i>Not identified^a</i>		-0.016	(0.024)
ATE	<i>Not identified^a</i>		-0.041*	(0.019)

Note: Average treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications.

* Statistical significance at 5% level.

^a Not identified due to insufficient common support.

These results suggest that it is important to reorient students to another track if they perform less well during secondary education to avoid high school drop-out. This finding is consistent with Dustmann et al. (2017), who find that the possibility to change tracks over time is important to mitigate negative long-term effects of early tracking.

Table 10: Downgrading: average marginal effects

	Coef.	St. error
Mother has no high school degree ^a	0.074*	(0.016)
Mother has high school degree ^a	0.047*	(0.011)
No Dutch at home	-0.049*	(0.015)
Low income	0.019*	(0.009)
Male	0.062*	(0.007)
Repeated	0.128*	(0.014)

Note: Average marginal effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. Average marginal effects of control variables report the change in the ATE when the value of the control variable would change from 0 to 1.

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

7 Conclusion

We study the causal and heterogeneous impact of choosing an elite high school on high school completion. Instead of offering study programs in different tracks, elite schools offer programs in the academic track only. We apply our analysis to an early tracking system in the region of Flanders in Belgium.

We find that there is substantial heterogeneity in the treatment effect, both within and between treatment groups. On average, students are 3.3 percentage points more likely to obtain a high school degree if they start at an elite school. However, students attending elite schools experience, on average, no effect of treatment, while students who do not choose an elite school would experience a significantly positive average effect of 5.1 percentage points. Part of this result is driven by greater treatment effects for students from disadvantaged backgrounds as they are less likely to choose an elite school.

Further analysis shows that this heterogeneity can be explained by tracking decisions. Elite schools incentivize non-treated students to enroll and perform well in the academic track, which increases their chances of obtaining a degree. Treated students are already more likely to choose the academic track in both types of schools but instead, have a higher risk of being mismatched because their high preference for the elite school interferes with their tracking decisions.

These results have important implications for educational policies. First, the average

positive effects of elite schools imply that more students and especially students from disadvantaged backgrounds should be encouraged to enroll in schools that offer only an academic track. More information about the capabilities of students to pursue an academic track could assist those students from disadvantaged backgrounds, in particular, to opt more often for elite high schools.

Second, the large heterogeneity in gains from elite and non-elite schools suggests that students may benefit considerably from school variety. Countries should not follow a one size fits all approach because allowing for different types of schools could improve degree completion. Students should have a choice between different schooling options within commuting distance. Countries that split students into different tracks through track-specific schools would benefit from also offering schools that offer multiple tracks.

Third, the reverse selection on gains shows that students are not optimally choosing their school. Thus, improved guidance or tests at the end of primary education may be useful. Furthermore, students with a high preference for elite schools but who are not able to pursue an academic track should be better guided towards other schools to prevent them from dropping out. Countries that organize their secondary education in track-specific schools could create clusters of different types of schools to smoothen the transition to other tracks if students do not perform well in the academic track.

Our analysis is limited to outcomes in secondary education. Graduating from elite schools could have benefits in higher education or on the labor market. Further research is needed to consider the effects of elite schools on other outcomes, such as success in higher education or on the labor market.

References

- Aakvik, A., Heckman, J. and Vytlacil, E. (2005), *Estimating Treatment Effects for Discrete Outcomes when Responses to Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs*, Journal of Econometrics 125 (1-2), 15-51
- Abdulkadiroğlu, A., Angrist, J. and Pathak, P. (2014), *The Elite Illusion: Achievement Effects at Boston and New York Exam Schools*, Econometrica, 82 (1), 137-196
- Altonji, J., Elder, T. and Taber, C. (2005), *An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling*, The Journal of Human Resources, 40 (4), 791-821
- Andresen, M. (2018), *Exploring marginal treatment effects: Flexible estimation using Stata*, The Stata Journal, 18 (1), 118-158
- Barrow, L., Whitmore Schanzenbach, D. and Claessens, A. (2015), *The Impact of Chicago's Small High School Initiative*, Journal of Urban Economics, 87, 100-113
- Basu, A., Heckman, J., Navarro-Lozano, S. and Urzua, S. (2007), *Use of Instrumental Variables in the Presence of Heterogeneity and Self-selection: An Application to Treatment of Breast Cancer Patients*, Health Economics, 16, 1133-1157
- Brodaty, T., Gary-Bobo, R. and Prieto, A. (2013), *Does Speed Signal Ability? The Impact of Grade Retention on Wages*, Working paper
- Card, D. (1995), *Using Geographic Variation in College Proximity to Estimate the Return to Schooling*, Aspects of Labour Economics: Essays in Honour of John Vanderkamp, University of Toronto Press
- Carneiro, P., Heckman, J., Vytlacil E. (2011), *Estimating Marginal Returns to Education*. American Economic Review, 101 (6), 2754-2781
- Carneiro, P., Lokshin, M. and Umapathi, N. (2016), *Average and Marginal Returns to Upper Secondary Schooling in Indonesia: Average and Marginal Returns to Schooling in Indonesia*. Journal of Applied Econometrics, 32 (1), 16-36
- Chmielewski, A. (2014), *An International Comparison of Achievement Inequality in Within- and Between-school Tracking Systems*, American Journal of Education 120 (3), 293-324

- Clark, D. and Del Bono, E. (2016), *The Long-run Effects of Attending an Elite School: Evidence from the UK*, American Economic Journal: Applied Economics, 8 (1), 150-176
- Cockx, B., Picchio, M. and Baert, S. (2018), *Modeling the Effects of Grade Retention in High School*, Journal of Applied Econometrics, 34 (3), 403-424
- Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. (2016), *From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions*, Labour Economics, 41, 47-60
- Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. (2018), *Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance*, Journal of Political Economy, 126 (6), 2356-2409
- Cummins, J. R. (2017), *Heterogeneous treatment effects in the low track: Revisiting the Kenyan primary school experiment*, Economics of Education Review, 56, 40–51
- De Groote, O. (2020), *A Dynamic Model of Effort Choice in High School*, TSE working paper - 1002
- Declercq, K. and Verboven, F. (2015), *Socio-economic Status and Enrollment in Higher Education: Do Costs Matter?*, Education Economics, 23 (5), 532-556
- Doyle, J. (2007), *Child Protection and Child Outcomes: Measuring the Effects of Foster Care*, American Economic Review, 97 (5), 1583-1610
- Doyle, J. (2008), *Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care*, Journal of Political Economy, 116 (4), 746-770
- 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, 1739–1774
- Dustmann, C., Puhani, P. and Schönberg, U. (2017), *The Long-term Effects of Early Track Choice*, Economic Journal, 127 (603), 1348-1380
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman & Hall.
- Fu, C. and Mehta, N. (2018), *Ability Tracking, School and Parental Effort, and Student Achievement: A Structural Model and Estimation*, Journal of Labor Economics, 36 (4), 923-979

- Galasso, A. and Schankerman, M. (2015), *Patents and Cumulative Innovation: Causal Evidence from the Courts*, *The Quarterly Journal of Economics*, 130 (1), 317-369
- Guyon, N., Maurin, E. and McNally, S. (2012), *The Effect of Tracking Students by Ability into Different Schools: A Natural Experiment*, *Journal of Human Resources*, 47 (3), 684-721
- Hanushek, E. and Woessmann, L. (2006), *Does Educational Tracking affect Performance and Inequality? Differences-in-differences Evidence across Countries*, *The Economic Journal*, 116, C63-C76
- Heckman, J. and Vytlacil, E. (1999), *Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects*, *Proceedings of the National Academy of Sciences*, 96 (8), 4730-4734
- Heckman, J. and Vytlacil, E. (2005), *Structural Equations, Treatment Effects, and Econometric Policy Evaluation*, *Econometrica*, 73 (3), 669-738
- Imbens, G. and Angrist, J. (1994), *Identification and Estimation of Local Average Treatment Effects*, *Econometrica*, 62 (2), 467-475
- Koop, G. and Poirier, Dale J. (1997), *Learning about the across-regime correlation in switching regression models*, *Journal of Econometrics*, 78, 217-227
- Lokshin, M. and Sajaia, Z. (2011), *Impact of Interventions on Discrete Outcomes: Maximum Likelihood Estimation of the Binary Choice Models with Binary Endogenous Regressors*, *The Stata Journal*, 11 (3), 368-385
- Manski, C., Sandefur G., McLanahan, S. and Powers, D. (1992), *Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation*, *Journal of the American Statistical Association*, 87 (417), 25-37
- Mühlenweg, A. M. and Puhani, P. A. (2010), *The Evolution of the School-Entry Age Effect in a School Tracking System*, *Journal of Human Resources*, 45, 407-438
- Nybom, M. (2017), *The Distribution of Lifetime Earnings Returns to College*, *Journal of Labor Economics*, 35 (4), 903-952
- OECD (2018), *Education at a Glance 2018: OECD Indicators*, OECD Publishing, Paris
- Oosterbeek, H., Ruijs, N. and de Wolf, I. (2020), *Using Admission Lotteries to Estimate Heterogeneous Effects of Elite Schools*, Tinbergen Institute Discussion Paper, TI 2020-018/V

- Pekkarinen, T., Uusitalo, R. and Kerr, S. (2009), *School Tracking and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform*, Journal of Public Economics, 93 (7-8), 965-973
- Pop-Eleches, C. and Urquiola, M. (2013), *Going to a Better School: Effects and Behavioral Responses*, American Economic Review, 103 (4), 1289-1324
- Poirier, Dale J. and Tobias, Justin, L. (2003), *On the Predictive Distributions of Outcome Gains in the Presence of an Unidentified Parameter*, Journal of Business & Economic Statistics, 21 (2), 258-268
- Roller, M. and Steinberg, D. (2020), *The distributional effects of early school stratification - non-parametric evidence from Germany*, European Economic Review, 125
- Vademecum Statistische sectoren (2012), http://statbel.fgov.be/nl/binaries/Secteurs%20stat-NL_tcm325-174181.pdf, [23/10/2015]
- VDAB (2019), *32ste schoolverlatersrapport - editie 2019*, <https://www.vdab.be/sites/web/files/doc/schoolverlaters/schoolverlatersrapport2019.pdf>
- Vytlacil, E. (2002), *Independence, Monotonicity, and Latent Index Models: An Equivalence Result*, Econometrica, 70 (1), 331-341

Appendix A: Construction of the dataset

We obtained our main dataset from the Flemish Ministry of Education. This dataset contains 72,292 and 72,581 students who started for the first time in the first year of secondary education in 2003 and 2004, respectively. We observe students in administrative records until 2012 but we do not use the final year for the first cohort so that we observe students in both cohorts during nine consecutive years. We keep only students who start at the comprehensive program. In the first year, students can also start at a pre-vocational program. Students starting at this program often did not successfully finish their secondary education and therefore could not enroll in elite schools. This reduces our sample to 60,754 students in 2003 and 60,992 students in 2004. We also remove double entries, reducing the sample by 26 observations. We remove four students who entered both an elite and a non-elite school. Next, we remove students who do not live in one of the Flemish provinces. This reduces our sample to 59,092 and 59,824 students, respectively.⁴⁰ Removing the observations with missing information about individual characteristics or detailed location reduces the sample to 55,527 and 56,252 observations, respectively.

We combine these data with data on neighborhood characteristics at the municipal and statistical sector level provided by ADSEI (now Statbel). From this data source, we use median income, average educational level, share of inhabitants with Belgian nationality, and population density. Most neighborhood characteristics are measured in 2001, before students enroll in secondary education. Median income is measured in 2004 because this is not available in the census of 2001. We merge the two datasets based on the statistical sector. A limited number of students cannot be merged and our sample reduces to 55,434 and 56,137 students for 2003 and 2004, respectively.

We observe for each student the study program in each year and the track as from grade 9. The track distinction does not officially exist before grade 9 but in grade 8, students already choose a specific study program that prepares them for a specific track (we confirmed this by looking at common switching patterns). In practice, tracking thus occurs as from grade 8. Based on the study program, we can assign each student to a specific study program from grade 8.

We observe the administrative number of the school and the exact location of the campus where the student is located. The administrative definition of a school does not always overlap with the actual school as it is perceived by parents and children. Large schools often

⁴⁰Note that the province of Brabant was split in three parts in 1995. We still use the historical province of Brabant, which means we include the students of the Flemish province of Flemish Brabant, but also the students in Brussels and Walloon Brabant who attend Flemish schools.

have several administrative entities at the same address. Some other schools use the same administrative entity for schools in very different locations. Therefore, we use the address of the campus where the student is located to create an identifier for each school. We assign students to elite and non-elite schools based on the tracks that students choose at these schools. We can compute travel distance to schools based on the exact location of the school and the centroid of the statistical sector where the student lives.

We construct our outcome variables as follows. First, a student graduates from high school if he or she obtains a degree within nine years of studying. Since all schools appear in our dataset, we assume drop out from the data is drop out from the schooling system. Second, a student graduates without a study delay if he or she obtains the degree within six years of studying. Finally, a student downgrades if he or she switches at least once to a lower ranked track.

Appendix B: Sensitivity analysis

In this appendix, we assess how sensitive our results are to the chosen sample of students, the composition of elite schools and the definition of obtaining a high school degree. The different treatment effects of choosing an elite school on high school completion are shown in Table B1. The first panel repeats the results from Table 5.

First, we restrict the sample to students for whom the exogeneity condition of the instrument is most likely to hold. Based on institutional grounds, we argue that distance is an exogenous instrument for school choice in Flanders when controlling for neighborhood characteristics. School choice is essentially free, and for most students, several schools are located within commuting distance. From our dataset, we have computed that for the median student, six schools are located within a 5 km distance. Therefore, it seems unlikely that parents would base their location decision on a preference for certain schools. If our instrument is not exogenous, we expect this violation of the exogeneity assumption to be stronger in areas with fewer school alternatives, because parents may then decide to live closer to their preferred school. We assess this issue and repeat the analysis on a subsample of the data where we include only students who have at least four schooling options, located within a 5 km radius. Sensitivity check A in Table B1 shows that the treatment effects derived from the parametric model are almost identical to the treatment effects on the whole sample. The ATT is statistically insignificant and the ATNT and ATE are significantly positive. However, the ATT derived from the semi-parametric model is significantly positive for this alternative sample.

In sensitivity check B, we restrict the sample to students living close to elite schools and

Table B1: Sensitivity analysis obtaining a high school degree: average treatment effects

	Semi-parametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error
Main outcome: Obtaining a high school degree (N=111,571)				
ATT	0.004	(0.019)	-0.016	(0.008)
ATNT	<i>Not identified^a</i>		0.051*	(0.010)
ATE	<i>Not identified^a</i>		0.033*	(0.008)
Sensitivity A: Students with at least four schools within 5 km (N=76,580)				
ATT	0.039*	(0.019)	-0.015	(0.011)
ATNT	<i>Not identified^a</i>		0.047*	(0.016)
ATE	<i>Not identified^a</i>		0.029*	(0.012)
Sensitivity B: Students with an elite school among the four closest schools (N=58,745)				
ATT	-0.028	(0.051)	-0.029	(0.022)
ATNT	<i>Not identified^a</i>		0.048	(0.037)
ATE	<i>Not identified^a</i>		0.020	(0.024)
Sensitivity C: Students who choose the academic track (N=73,066)				
ATT	0.026	(0.015)	-0.008	(0.007)
ATNT	<i>Not identified^a</i>		0.028*	(0.007)
ATE	<i>Not identified^a</i>		0.014*	(0.005)
Sensitivity D: Students in schools that offer the academic track (N=97,297)				
ATT	0.014	(0.019)	-0.012	(0.007)
ATNT	<i>Not identified^a</i>		0.046*	(0.010)
ATE	<i>Not identified^a</i>		0.028*	(0.007)
Sensitivity E: Students still in school after three years of study delay omitted (N=111,275)				
ATT	-0.001	(0.018)	-0.018*	(0.008)
ATNT	<i>Not identified^a</i>		0.049*	(0.010)
ATE	<i>Not identified^a</i>		0.031*	(0.008)
Sensitivity F: Students still in school after three years of study delay classified as graduated (N=111,571)				
ATT	-0.002	(0.018)	-0.018*	(0.008)
ATNT	<i>Not identified^a</i>		0.049*	(0.011)
ATE	<i>Not identified^a</i>		0.031*	(0.008)

Note: Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. 3

* Statistical significance at 5% level.

^a Not identified due to insufficient common support.

consider only students who have at least one elite school among the four closest schooling options. If parents base their location decision on the availability of elite schools and move to an area with an elite school, we would expect that the exogeneity condition of the instrument is least likely to hold for students living in neighborhoods close to these schools. Similar to our main specification, we obtain a statistically insignificant ATT in both approaches. The ATE and ATNT derived from the parametric model are similar to the estimates from the main specification, but less precisely estimated due to the smaller sample size.

In sensitivity check C, we restrict the sample to students starting the academic track in the second grade of high school. This removes a large number of students, primarily in the non-elite schools. Many of them are possibly not interested in staying in an elite school if they would go there in the first year, as they are not willing to enroll in the academic track. Therefore, it could generate more precise estimates by focusing on a more relevant group of students. At the same time, it also removes some students who did want to enroll in the academic track before they made their school choice but decided not to (or were not permitted to). Note that we found that the elite school itself has a positive impact on choosing the academic track for non-treated students and this improves their outcome. Therefore, we should expect a less positive effect by removing these students as we are ignoring part of the effect of an elite school. Indeed, the effects for the non-treated are smaller but more precisely estimated in this restricted sample.

In sensitivity check D, we assess whether or not our results are robust to a different composition of non-elite schools. There exist two different types of non-elite schools. The first type offer programs in the academic track in combination with technical, artistic or vocational programs. Most first-year students (60.5%) enroll at this type of school. The second type of schools specialize in technical, artistic and/or vocational programs and do not offer academic programs. These schools have a small share of first-year enrollment (12.8%). As students who start at a non-elite school without programs in the academic track can differ from students who start at the first type of non-elite schools, we repeat the analysis and include only students who started at an elite school and students who started at a non-elite school that also offers the academic track. We obtain similar results as in our main specification.

In the last two panels of Table B1, we assess whether or not our results are affected by our assumptions on high school graduation. We observe students for a maximum of nine years in high school: 0.3% of the students have not obtained a degree at the end of the ninth year and are still enrolled in high school. As our main outcome variable is high school completion, we considered these students as high school drop-outs in our previous analysis. We assess whether or not our results are affected by this assumption and remove these

students from the data in sensitivity check E. In sensitivity check F, we assume that these students eventually obtain a high school degree and consider them as high school graduates. The last two panels in Table B1 show that our results do not change when imposing other assumptions on the study outcomes of these students.

Appendix C: Role of control variables

In this appendix, we assess how including neighborhood characteristics affects the treatment effects of choosing an elite school on high school completion. Table C1 summarizes the treatment effects based on the parametric model for several specifications with alternative sets of control variables. When including only student characteristics we find an ATT of -4.6 percentage points that is statistically significantly different from zero. The ATE and ATNT are precisely estimated around 0. The effect of an elite school improves substantially when we add a small set of dummy variables that capture the different locations of students: four categories of urbanization of the municipality and five dummy variables for provinces. The ATT is now only -1.9 percentage points and the ATE and ATNT become statistically significant, with positive treatment effects of 2.3 and 3.9 percentage points. Adding additional characteristics at the municipality level hardly affects the ATT and slightly increases the ATE and ATNT. In the last specification (which is the one we use in the paper) we also control for demographics at the statistical sector level which changes little to the results.

In Table C2 we show the impact of student characteristics on relative distance when controlling for different sets of regional characteristics. The first specification shows that individual characteristics explain only a small part of the variation ($R^2=0.006$). Effect sizes are relatively small too, but the large sample size does allow us to detect statistically significant effects that can be informative to learn about a potential bias in the estimates when we do not control for neighborhood characteristics. Importantly, we see that students who repeated a grade before entering secondary education live 263 meters closer to an elite school. This might seem like a small distance, but it is statistically significant. We interpret this as an indication that students of lower ability live closer to elite schools. Since this dummy variable is only a very rough proxy of ability, controlling for it is not sufficient to obtain reliable estimates of treatment effects. After controlling for province and urbanization fixed effects (FE), we explain much more of the variation ($R^2=0.142$) and the effect of grade repetition on the instrument reduces to a very small and statistically insignificant 10 meters. This suggests that we are now better controlling for local differences in ability. Consistent with this, we observe a strong improvement in the estimated effect of elite schools in Table C1. Other controls for neighborhood characteristics help to further reduce the correlation

between socioeconomic characteristics and the instrument, but this has much smaller effects on the estimates of the treatment effects.

We conclude that Flanders is not sufficiently homogeneous to rely on distance as an instrument for school choice without controlling for the broad differences between provinces and levels of urbanization. Descriptive statistics show that proportionally more students of lower ability are living in urban areas and that elite schools are more accessible for students living in these areas. For example, 93% of the 75,595 students living in the most rural municipalities graduate from high school while this is only 86% of the 15,808 students living in the largest cities. While 9% of students repeated a grade before entering secondary education in rural areas, 19% of students were retained in the largest cities. Meanwhile, the average difference in distance to an elite school is only 835 meters for students living in the largest cities and 3013 meters for students living in the most rural municipalities. This is not surprising, as a larger number of potential students in urban municipalities allows for more variety in school choice and the possibility for schools to specialize in a track. It is therefore much easier to access an elite school for students living in urban areas, explaining why 38% of them go to an elite school. In rural municipalities, this is only 24%. Without including controls for the different types of municipalities, the estimator links lower performance in urban municipalities (which are likely due to other factors than elite schools) with easier access to elite schools, leading to an underestimation of the treatment effects.

Table C1: The role of control variables: average treatment effects

	(1)		(2)		(3)		(4)	
	Coef.	St. error	Coef.	St. error	Coef.	St. error	Coef.	St. error
ATT	-0.046*	(0.004)	-0.019*	(0.008)	-0.015	(0.009)	-0.016	(0.008)
ATNT	0.005	(0.014)	0.039*	(0.010)	0.047*	(0.010)	0.051*	(0.011)
ATE	-0.009	(0.011)	0.023*	(0.008)	0.030*	(0.008)	0.033*	(0.009)
Student characteristics	yes		yes		yes		yes	
Province and urbanization FE	no		yes		yes		yes	
Neighborhood characteristics	no		no		municipality		municipality + statistical sector	

Note: Average treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed by sampling 250 draws from the estimated distribution. For each draw of the estimates, we calculate the treatment effects by simulating the error terms of the potential outcome framework (100 draws for each student).

* Statistical significance at 5% level.

Table C2: The role of control variables: regression of relative distance on student characteristics

Variables	(1)		(2)		(3)		(4)	
	Coef.	St. error	Coef.	St. error	Coef.	St. error	Coef.	St. error
Mother no high school degree ^a	-0.437*	(0.047)	-0.349*	(0.041)	-0.101*	(0.038)	-0.055	(0.033)
Mother high school degree ^a	-0.370*	(0.033)	-0.224*	(0.030)	-0.055*	(0.027)	-0.023	(0.024)
No Dutch at home	0.803*	(0.083)	0.158*	(0.069)	0.092	(0.068)	0.040	(0.062)
Low income	-0.135*	(0.037)	-0.124*	(0.032)	-0.062*	(0.031)	-0.046	(0.030)
Male	-0.009	(0.021)	-0.006	(0.019)	-0.012	(0.019)	-0.010	(0.019)
Repeated	0.263*	(0.039)	0.010	(0.034)	-0.028	(0.032)	-0.025	(0.032)
Province and urbanization FE	no		yes		yes		yes	
Neighborhood characteristics	no		no		municipality		municipality + statistical sector	
F-test individual characteristics	49.81 (p=0.000)		16.60 (p=0.000)		2.52 (p=0.0195)		1.28 (p=0.263)	
R ²	0.006		0.142		0.183		0.185	
Observations	111,571		111,571		111,571		111,571	

Note: Standard errors are corrected for clustering within the statistical sector.

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Appendix D: Additional tables and figures

Table D1: Obtaining a high school degree: OLS and 2SLS

Variables	OLS				2SLS			
	(1)		(2)		(3)		(4)	
	Coef.	St. error						
Elite school	0.039*	(0.002)	0.018*	(0.002)	-0.005	(0.010)	-0.009	(0.010)
Mother no high school ^a			-0.086*	(0.003)			-0.091*	(0.003)
Mother high school ^a			-0.032*	(0.002)			-0.036*	(0.002)
No Dutch at home			-0.046*	(0.006)			-0.044*	(0.006)
Low income			-0.001	(0.002)			-0.002	(0.002)
Male			-0.048*	(0.002)			-0.048*	(0.002)
Repeated			-0.097*	(0.004)			-0.101*	(0.004)
Constant	0.730*	(0.028)	0.902*	(0.027)	0.770*	(0.030)	0.932*	(0.029)
Neighborhood characteristics	yes		yes		yes		yes	
Observations	111,571		111,571		111,571		111,571	

Note: Standard errors are corrected for clustering within the statistical sector. Neighborhood characteristics include population density, median household income, the share of Belgians and average educational level at the municipality and the statistical sector. All regressions additionally control for province fixed effects and indicators for the level of urbanization of the municipality.

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Figure D1: Weights for the calculation of the ATT

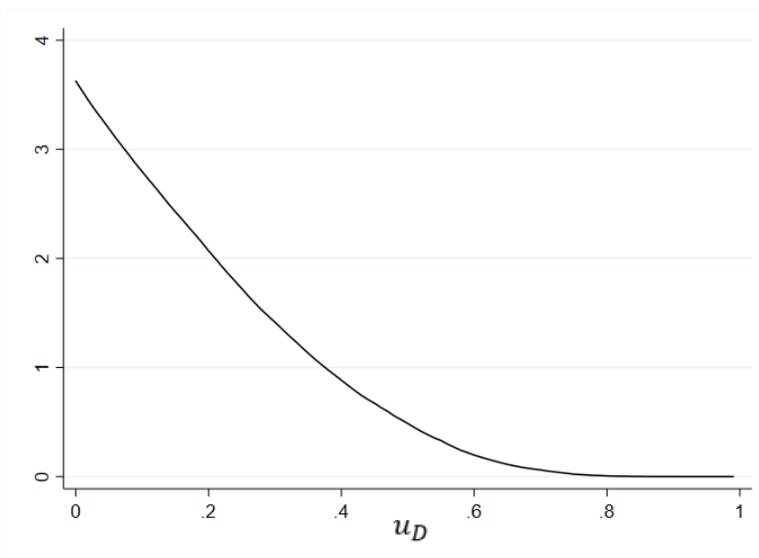


Table D2: Obtaining a high school degree: estimation results semi-parametric approach

Variables	Selection equation		Outcome equation			
	Index prop. score		Baseline effect		Difference if treated	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Relative distance	0.131*	(0.004)				
Mother no high school ^a	-0.659*	(0.015)	-0.087*	(0.005)	0.026	(0.024)
Mother high school ^a	-0.406*	(0.010)	-0.030*	(0.003)	-0.008	(0.011)
No Dutch at home	0.251*	(0.025)	-0.025*	(0.010)	-0.075*	(0.024)
Low income	-0.119*	(0.013)	0.006	(0.004)	-0.030	(0.016)
Male	-0.042*	(0.010)	-0.056*	(0.003)	0.030*	(0.008)
Repeated	-0.496*	(0.018)	-0.100*	(0.006)	0.050	(0.029)
Neighborhood characteristics		yes		yes		yes
P-value no observable heterogeneity		0.000				
P-value no unobservable heterogeneity		0.424				
Observations		111,571				

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a semi-parametric approach that follows the approach described in Section 4.3 of Cornelissen et al. (2016) (pp. 55-56). The estimates of “Baseline effect” correspond to β_0 in their equation (26), while “Difference if treated” corresponds to $\beta_1 - \beta_0$. The reported test statistic for observable heterogeneity tests for the joint significance of all elements in the vector $\beta_1 - \beta_0$, while the test for unobservable heterogeneity tests if MTEs differ with unobserved costs of treatment. We estimated the model in STATA with the user-written command "mtefe" (Andresen, 2018).

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Table D3: Obtaining a high school degree: estimation results parametric approach

Variables	Selection equation		Outcome equations			
	elite school		elite school		non-elite school	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Relative distance	0.131*	(0.004)				
Mother no high school ^a	-0.659*	(0.015)	-0.457*	(0.051)	-0.632*	(0.021)
Mother high school ^a	-0.406*	(0.010)	-0.293*	(0.035)	-0.326*	(0.018)
No Dutch at home	0.253*	(0.025)	-0.290*	(0.048)	-0.131*	(0.029)
Low income	-0.118*	(0.013)	-0.064	(0.035)	-0.004	(0.016)
Male	-0.042*	(0.010)	-0.413*	(0.027)	-0.343*	(0.013)
Repeated	-0.496*	(0.018)	-0.458*	(0.051)	-0.473*	(0.018)
Constant	2.388*	(0.261)	1.831*	(0.417)	1.844*	(0.222)
Neighborhood characteristics		yes		yes		yes
ρ_1	-0.185*	(0.070)				
ρ_0	0.193*	(0.058)				
Log likelihood		-82068				
Observations		111,571				

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a parametric model where we assume that the error terms are jointly normally distributed. We estimated the model in STATA with the user-written command "switch_probit" (Lokshin and Sajaia, 2011).

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Table D4: Graduating without study delay: average treatment effects

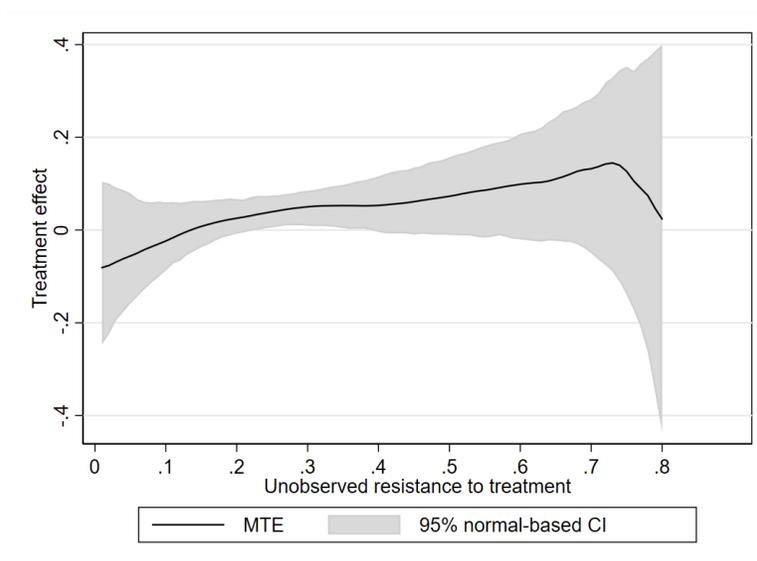
	Semi-parametric approach		Parametric approach	
	Coef.	St. error	Coef.	St. error
ATT	0.016	(0.030)	-0.010	(0.018)
ATNT	<i>Not identified^a</i>		0.112*	(0.020)
ATE	<i>Not identified^a</i>		0.080*	(0.016)

Note: Average treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed with a bootstrap procedure using 250 replications. Average outcome in sample: 0.775 (Elite), 0.700 (non-elite).

* Statistical significance at 5% level.

^a Not identified due to insufficient common support.

Figure D2: Graduating without study delay: MTE



Note: MTEs are calculated based on 111,571 students using a local polynomial regression of degree 2 with an Epanechnikov kernel. Standard errors are computed with a bootstrap procedure (250 replications) and clustered within the statistical sector. P-value no observable heterogeneity (0.000), p-value no unobservable heterogeneity (0.882).

Table D5: Graduating without study delay: estimation results semi-parametric approach

Variables	Selection equation		Outcome equation			
	Index prop. score		Baseline effect		Difference if treated	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Relative distance	0.131*	(0.004)				
Mother no high school ^a	-0.659*	(0.015)	-0.162*	(0.007)	-0.004	(0.035)
Mother high school ^a	-0.406*	(0.010)	-0.082*	(0.006)	-0.040	(0.021)
No Dutch at home	0.251*	(0.025)	-0.093*	(0.013)	-0.011	(0.034)
Low income	-0.119*	(0.013)	-0.009	(0.006)	-0.055*	(0.021)
Male	-0.042*	(0.010)	-0.149*	(0.004)	0.162	(0.013)
Repeated	-0.496*	(0.018)	-0.042*	(0.007)	-0.035	(0.035)
Neighborhood characteristics		yes		yes		yes
P-value no observable heterogeneity		0.000				
P-value no unobservable heterogeneity		0.882				
Observations		111,571				

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a semi-parametric approach that follows the approach described in Section 4.3 of Cornelissen et al. (2016) (pp. 55-56). The estimates of “Baseline effect” correspond to β_0 in their equation (26), while “Difference if treated” corresponds to $\beta_1 - \beta_0$. The reported test statistic for observable heterogeneity tests for the joint significance of all elements in the vector $\beta_1 - \beta_0$, while the test for unobservable heterogeneity tests if MTEs differ with unobserved costs of treatment. We estimated the model in STATA with the user-written command "mtefe" (Andresen, 2018).

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Table D6: Graduating without study delay: estimation results parametric approach

Variables	Selection equation		Outcome equations			
	elite school		elite school		non-elite school	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Relative distance	0.131*	(0.004)				
Mother no high school ^a	-0.659*	(0.015)	-0.512*	(0.037)	-0.539*	(0.159)
Mother high school ^a	-0.405*	(0.011)	-0.318*	(0.024)	-0.318*	(0.013)
No Dutch at home	0.251*	(0.025)	-0.307*	(0.039)	-0.227*	(0.025)
Low income	-0.118*	(0.013)	-0.122*	(0.025)	-0.054*	(0.012)
Male	-0.042*	(0.010)	-0.501*	(0.018)	-0.454*	(0.010)
Repeated	-0.496*	(0.018)	-0.212*	(0.040)	-0.146*	(0.016)
Constant	2.384*	(0.261)	0.894*	(0.287)	0.788*	(0.175)
Neighborhood characteristics		yes		yes		yes
ρ_1	-0.172*	(0.047)				
ρ_0	0.137*	(0.037)				
Log likelihood	-114367					
Observations	111,571					

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a parametric model where we assume that the error terms are jointly normally distributed. We estimated the model in STATA with the user-written command "switch_probit" (Lokshin and Sajaia, 2011).

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Table D7: Extended model: estimation results selection equations

Variables	Selection equation elite school		Selection equation academic track			
	Coef.	St. error	after non-elite		after elite	
			Coef.	St. error	Coef.	St. error
Relative distance elite school	0.206*	(0.006)				
Relative distance academic track	-0.103*	(0.014)	0.032*	(0.011)	-0.015	(0.016)
Mother no high school ^a	-0.931*	(0.021)	-1.268*	(0.042)	-0.889*	(0.085)
Mother high school ^a	-0.573*	(0.015)	-0.879*	(0.030)	-0.524*	(0.054)
No Dutch at home	0.356*	(0.036)	0.559*	(0.039)	-0.012	(0.059)
Low income	-0.164*	(0.018)	-0.135*	(0.016)	-0.163*	(0.037)
Male	-0.063*	(0.014)	-0.309*	(0.014)	-0.220*	(0.027)
Repeated	-0.693*	(0.025)	-1.192*	(0.041)	-0.839*	(0.071)
Factor loading	1	(.)	0.870*	(0.058)	0.251	(0.156)
Neighborhood characteristics		yes		yes		yes
Log likelihood		-137321				
Observations		111,571				

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a factor model. We estimated the model with the STATA command “gsem”. Effects of neighborhood characteristics on track choice are constrained to be the same, regardless of elite school choice.

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Table D8: Extended model: estimation results outcome equations

Variables	Outcome equations							
	Academic track				Other track			
	elite school		non-elite school		elite school		non-elite school	
	Coef.	St. error	Coef.	St. error	Coef.	St. error	Coef.	St. error
Relative distance elite school								
Relative distance academic track								
Mother no high school ^a	-0.379*	(0.052)	-0.537*	(0.054)	-0.224	(0.119)	-0.558*	(0.071)
Mother high school ^a	-0.254*	(0.035)	-0.302*	(0.039)	-0.143	(0.097)	-0.244*	(0.052)
No Dutch at home	-0.367*	(0.056)	-0.181*	(0.046)	-0.097	(0.119)	-0.138*	(0.043)
Low income	-0.071	(0.040)	-0.086*	(0.027)	0.077	(0.082)	0.050*	(0.021)
Male	-0.412*	(0.030)	-0.438*	(0.024)	-0.392*	(0.073)	-0.269*	(0.023)
Repeated	-0.394*	(0.055)	-0.497*	(0.055)	-0.207*	(0.100)	-0.395*	(0.053)
Factor loading	-0.247*	(0.116)	0.147	(0.087)	-0.448*	(0.150)	0.186	(0.129)
Neighborhood characteristics		yes		yes		yes		yes
Log likelihood	-137321							
Observations	111,571							

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a factor model. We estimate the model with the STATA command “gsem”. Effects of neighborhood characteristics on the outcome are constrained to be the same for each track choice but are allowed to differ by school choice.

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Table D9: Extended model using track in grade 9: average treatment effects

	Coef.	St. error	Coef.	St. error	Coef.	St. error
Treatment	Elite school		Academic track		Elite school	
Outcome	High school degree		High school degree		Academic track	
ATT	-0.010	(0.008)	0.072*	(0.009)	0.070*	(0.011)
ATNT	0.048*	(0.010)	0.069*	(0.010)	0.175*	(0.030)
ATE	0.032*	(0.008)	0.072*	(0.008)	0.147*	(0.024)

Note: Average treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed by sampling 250 draws from the estimated distribution. For each draw of the estimates, we calculate the treatment effects by simulating the error terms of the potential outcome framework (100 draws for each student).

* Statistical significance at 5% level.

Table D10: Extended model using track in grade 9: average treatment effects conditional upon track choice

	No restrictions		Only academic track		No academic track	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
ATT	-0.010	(0.008)	-0.003	(0.006)	-0.072*	(0.023)
ATNT	0.048*	(0.010)	0.037*	(0.012)	0.027	(0.015)
ATE	0.032*	(0.008)	0.026*	(0.009)	0.000	(0.013)

Note: Average treatment effects are calculated based on 111,571 students. Standard errors are corrected for clustering within the statistical sector and computed by sampling 250 draws from the estimated distribution. For each draw of the estimates we calculate the treatment effects by simulating the error terms of the potential outcome framework (100 draws for each student).

* Statistical significance at 5% level.

Table D11: Downgrading: estimation results semi-parametric approach

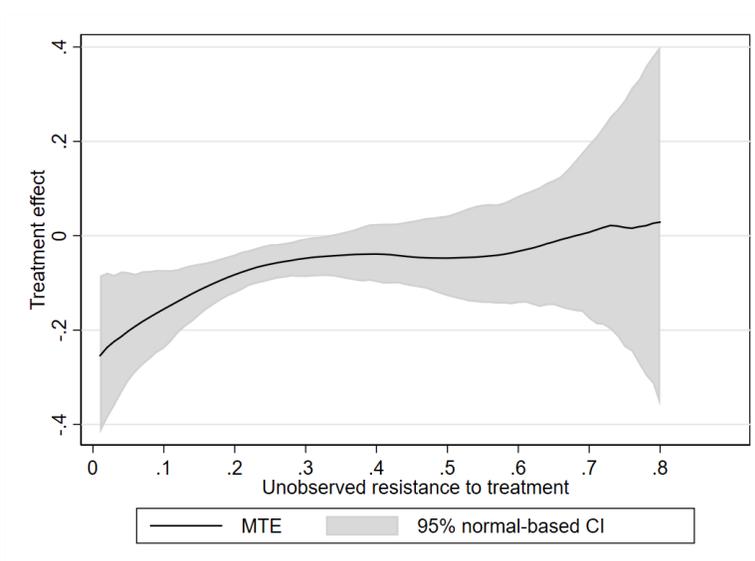
Variables	Selection equation		Outcome equation			
	Index prop. score		Baseline effect		Difference if treated	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Relative distance	0.131*	(0.004)				
Mother no high school ^a	-0.659*	(0.015)	0.128*	(0.008)	0.126*	(0.036)
Mother high school ^a	-0.406*	(0.010)	0.100*	(0.007)	0.049*	(0.023)
No Dutch at home	0.251*	(0.025)	-0.013	(0.013)	0.001	(0.034)
Low income	-0.119*	(0.013)	0.014*	(0.006)	0.040	(0.023)
Male	-0.042*	(0.010)	0.024*	(0.005)	0.099*	(0.014)
Repeated	-0.496*	(0.018)	0.023*	(0.008)	0.240*	(0.039)
Neighborhood characteristics		yes		yes		yes
P-value no observable heterogeneity		0.000				
P-value no unobservable heterogeneity		0.913				
Observations		111,571				

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a semi-parametric approach that follows the approach described in Section 4.3 of Cornelissen et al. (2016) (pp. 55-56). The estimates of “Baseline effect” correspond to β_0 in their equation (26), while “Difference if treated” corresponds to $\beta_1 - \beta_0$. The reported test statistic for observable heterogeneity tests for the joint significance of all elements in the vector $\beta_1 - \beta_0$, while the test for unobservable heterogeneity tests if MTEs differ with unobserved costs of treatment. We estimated the model in STATA with the user-written command "mtefe" (Andresen, 2018).

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.

Figure D3: Downgrading: MTE



Note: MTEs are calculated based on 111,571 students using a local polynomial regression of degree 2 with an Epanechnikov kernel. Standard errors are computed with a bootstrap procedure (250 replications) and clustered within the statistical sector. P-value no observable heterogeneity (0.000), p-value no unobservable heterogeneity (0.913).

Table D12: Downgrading: parametric approach

Variables	Selection equation		Outcome equations			
	elite school		elite school		non-elite school	
	Coef.	St. error	Coef.	St. error	Coef.	St. error
Relative distance	0.131*	(0.004)				
Mother no high school ^a	-0.659*	(0.015)	0.628*	(0.033)	0.378*	(0.016)
Mother high school ^a	-0.406*	(0.010)	0.468*	(0.021)	0.291*	(0.013)
No Dutch at home	0.251*	(0.025)	-0.143*	(0.040)	0.007	(0.023)
Low income	-0.119*	(0.013)	0.107*	(0.023)	0.046*	(0.012)
Male	-0.043*	(0.010)	0.296*	(0.016)	0.101*	(0.009)
Repeated	-0.497*	(0.018)	0.467*	(0.036)	0.106*	(0.015)
Constant	2.394*	(0.261)	-1.744*	(0.275)	-0.735*	(0.159)
Neighborhood characteristics		yes		yes		yes
ρ_1	-0.006	(0.040)				
ρ_0	0.123*	(0.034)				
Log likelihood	-121935					
Observations	111,571					

Note: Standard errors are corrected for clustering within the statistical sector. The results are obtained from a parametric model where we assume that the error terms are jointly normally distributed. We estimated the model in STATA with the user-written command "switch_probit" (Lokshin and Sajaia, 2011).

* Statistical significance at 5% level.

^a Base category = mother has a degree in higher education.