The choice of academic effort in high school.

Olivier De Groote*

October 31, 2018

Abstract

I estimate a dynamic model of educational decisions in high school that allows for observed and unobserved differences in initial ability. Each year, students choose their level of effort by choosing the academic level of their study program and by setting their distribution of end-of-year performance. Good performance is costly, but necessary to continue in the program and to graduate. This replaces traditional approaches, which assume performance follows an exogenous law of motion. I use the model to investigate tracking policies that aim to match students to the right program and obtain the following main findings: (1) encouraging underperforming students to switch to less academic programs substantially reduces grade retention and drop out, (2) there is no effect on obtaining a higher education degree, and (3) a model that assumes performance is exogenous would lead to large biases in many predicted outcomes and would falsely conclude that graduation rates in higher education are negatively affected by the policy change because it ignores increases in unobserved study effort.

Keywords: high school curriculum, early tracking, dynamic discrete choice, CCP estimation

JEL: C61, I26, I28

*Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France. E-mail: olivier.de-groote@tse-fr.eu. This paper benefited from helpful comments at various stages from Peter Arcidiacono, Estelle Cantillon, Jan De Loecker, Koen Declercq, Arnaud Maurel, Jo Van Biesebroeck and Frank Verboven and several audiences at KU Leuven and Duke University. It also benefited from discussions during and after presentations at ULiège, RUG Groningen, UAB Barcelona, Toulouse School of Economics, University of Cambridge, LMU Munich, McGill University, UNC Chapel Hill, Oslo University, ENSAE-CREST, IWAEE 2017, EALE 2017, ECORES and LEER summer schools 2017, NESG 2018 and the CEPR Applied IO 2018. I also thank Jan Van Damme, Bieke De Fraine and the Flemish Ministry of Education and Training, for providing the LOSO dataset. Funding for this project was generously provided by the Research Foundation Flanders (FWO).

1 Introduction

Students follow different curricula during secondary education, depending on their preferences and ability. Many countries separate students in academic or vocational tracks. Academic curricula do not focus on skills that are directly useful on the labor market but provide preparation for higher education. To achieve the European 2020 target of 40% college educated people, many countries aim to induce more students to choose academic curricula. Other countries often provide course-level differentiation. In the US, there is a similar trend towards more academic course taking, especially in STEM (Science, Technology, Engineering, Math)-fields.[1]

This trend raises two related concerns. First, it is unclear whether there is a causal effect of a more academic curriculum on success in higher education.[2] Second, not every student is expected to gain from an academic program. Students who are unlikely to go to college would waste time and effort they could otherwise spend on training skills that are of direct use on the labor market. They might also not have the required academic ability to finish the program successfully. Mismatch and failure can lead to unfavorable outcomes like grade retention and drop out. These outcomes do not only generate large costs for students, but also cause negative externalities on society. I therefore investigate how to design policies that help in matching students to a study program. This is especially a concern in early tracking countries, i.e. countries that differentiate students at the age of 10 to 12.[3]

To investigate the impact of high school curriculum and the design of suitable policies, I use a dataset that combines data on study program attendance and performance in secondary education with data on higher education. I use rich micro-data of Flanders, the largest region of Belgium. As in many countries, study programs consist of tracks and elective courses within each track. Students choose a program at age 12 but can update their choice almost every year after. There is a tracking policy that offers underperforming students the choice to switch out of an academically rigorous program or repeat the grade. First, I study the impact of study programs that differ in their academic level on long run outcomes and the extent to which the current tracking policy helps to improve these outcomes. Next, I look at an alternative policy that aims to minimize grade retention, at the cost of graduation rates of programs that prepare for higher education.

I develop a dynamic model of educational decisions in which high school students make

[1]The 2011 NAEP report compares high school students graduating in 2005 to students graduating in 1990. They find that they take more academic credits (16 on average instead of 13.7). The percentage of students that followed a rigorous curriculum also increased from 5% to 13% Nord et al. (2011).

[2]In the rest of this paper I will refer to study programs as the curriculum a student follows during secondary education.

[3]Germany and Austria already differentiate from the age of 10. Belgium and the Netherlands differentiate from age 12. Most of these early tracking countries also face much higher rates of grade retention OECD (2013).

yearly decisions about their level of academic effort. They do this by choosing among study programs of different level and by setting the distribution of their end-of-year performance. Allowing students to influence both the study program and the distribution of performance is novel and it is particularly important for the counterfactual simulations in this paper as we expect students to change their (unobserved) study effort in response to tracking policies. The decision of a study program is based on the (psychic) effort cost of studying today and the impact on future utility. The effort cost depends on (1) a fixed cost, independent of the expected performance, and (2) a variable cost, increasing in the probability to have a good performance outcome at the end of the year. Good performance is costly, but required to continue in the program and, at the end of high school, to graduate. Attending a high-level study program is expected to improve higher education outcomes. To identify these effects, I simultaneously estimate the causal effect of high school study programs on enrollment in and graduation from higher education. Because of unobserved ability, it is crucial to allow for correlation between the unobservables that impact higher education and choices in secondary education. I model this using a finite mixture of types with each type having different effort costs and higher education outcomes. Rich panel data and exclusion restrictions help in identifying the types without relying on arbitrary functional form assumptions. The Flemish context is particularly useful for this purpose as students make choices and obtain performance outcomes during many years. The extensive data on initial ability and socioeconomic status also allows me to include rich patterns of heterogeneity, even with a small number of unobserved types.

This paper contributes to three strands of literature. A first strand of literature investigates the causal impact of high school curriculum on long run educational and labor market outcomes.[4] Altonji (1995) finds small effects for several high school courses in the US but specifically states the difficulties in estimating causal effects are too large to draw policy conclusions on the results. Several papers look at the impact of intensive math courses and found positive effects, at least for some groups of students (Rose and Betts (2004), Joensen and Nielsen (2009), Aughinbaugh (2012)). Papers that look at choices between academic and vocational courses stress the importance of comparative advantages in different programs which causes heterogeneous effects (Kreisman and Stange (2017), Meer (2007)). I contribute to this literature by estimating the causal impact of multiple high school programs on higher education enrollment and graduation. I distinguish between four tracks that differ in their academic level and look at differences within tracks that prepare for higher education, based on the math-intensity of the curriculum and if classical languages are included. The benefit of the Flemish institutional context is that the study program does not have an impact on

[4]See also the review of Altonji *et al.* (2012).

the higher education options students can choose from. This allows me to identify the effect of both academic and vocational study programs. I also estimate a model that explains how students transition from one program to another during secondary education. This allows me to investigate policies that help to match students to a program and assess how it influences grade retention, drop out and higher education outcomes.

A second strand of literature looks at the impact of tracking policies during secondary education. Most papers look at the age in which students are separated into different tracks (see e.g. Hanushek and Woessman (2006), Pekkarinen *et al.* (2009)) or the long-run impact of the academic track for marginal students or students who are affected by specific policies (Guyon *et al.* (2012), Dustmann *et al.* (2017)). Baert *et al.* (2015) also look at the impact on students of being forced to switch track but only investigate outcomes during secondary education and do not compare different policies. Recent evidence also shows that switching track can diminish negative consequences of early track choice, suggesting choices during secondary education are important to further investigate (see De Groote and Declercq (2018) and Dustmann *et al.* (2017)). I contribute to this literature by investigating how tracking policies during secondary education can help underperforming students to switch to the right track.

Finally, this paper contributes to the development and estimation of dynamic discrete choice models in general, and educational decision making in particular. After the seminal contribution of Keane and Wolpin (1997), dynamic discrete choice models have often been used to evaluate the impact of counterfactual policies on choices of study programs or educational attainment.[5] Dynamics are important because students are expected to react in advance to changes in policy that would only affect them in the future. The only channel through which students can respond to changes in the future is through discrete, observable choices like college major (study program) or years of schooling. This excludes any response through the study effort they exert during the year. Theoretical and reduced form evidence suggests that dynamic incentives should also impact performance directly as students will change their study effort (Costrell (1994), Dubois *et al.* (2012), Garibaldi *et al.* (2012)). Recent papers have therefore included observable measures of effort in the model (Todd and Wolpin (2018), Mehta and Fu (2018)), but these are often unavailable and can only proxy for the actual effort students exert. I therefore propose an alternative approach that can be used when only program and performance data are available by rationalizing the observed distribution of performance as the result of an optimization process.

[5] See e.g. Eckstein and Wolpin (1999), Arcidiacono (2005), Joensen and Mattana (2017) or Declercq and Verboven (2018). The same type of models have also been used to look at the impact of wage returns on program choices (Arcidiacono (2004), Beffy *et al.* (2012).).

Performance often enters the model as a law of motion that characterizes the state variables in the next period. This law of motion is assumed to be exogenous, after conditioning on the program choice. I change the model by adding a choice variable that allows students to set the distribution of performance directly. Identification does not require additional data (like a measure of study effort) or exclusion restrictions. Instead, I make use of an economic condition that equates the (unobserved) marginal cost of improving the distribution of performance to its (observed) expected marginal benefit it will generate in the future. I also show that the model can be estimated without solving it by applying CCP (Conditional Choice Probability) estimation (Hotz and Miller (1993), Arcidiacono and Miller (2011)). The added choice of the performance distribution offers an additional reason to use CCP estimation as we not only avoid solving for optimal program choices, but also the optimal choice of the performance distribution. This can also be useful in other applications of dynamic discrete choice models where agents are expected to have a direct (but costly) impact on the distribution of state variables.

I find that studying an academic program in high school has a positive impact on enrollment in and graduation from higher education. At the same time, grade retention in high school decreases the probability to graduate from higher education. In the evaluation of different tracking policies, I find that allowing underperforming students to switch tracks as an alternative for repeating a grade has important benefits in the long run. Without this, the percentage of students with grade retention would increase by 8.1 %points, and college graduation would decrease by 1.8 %points. This suggests that underperforming students should be encouraged to switch to a program of lower academic level, rather than spending extra time in school to graduate from a more academic program. I also find that this policy can be further improved to avoid costly grade retention, without decreasing graduation from higher education. Prohibiting students to repeat a grade if they can avoid this by switching programs would decrease the number of students who were retained in secondary education by 10.7 %points (or 1 out of 3 students) and drop out by 1.4 %points.

Finally, I find that a model where the distribution of performance is exogenous, underestimates the positive effects on student outcomes of both counterfactual simulations of the tracking policy. This is because performance of students improves, possibly through an increase in unobserved study effort. The increase in drop out rates in a policy that forces underperforming students to repeat grades is 5 %points instead of 3, and the decrease in the policy where students are encouraged to switch to a different program is 1.1 %points instead of 1.4. We also see large differences in grade retention and for graduation from higher education we would falsely conclude that there would be a negative impact for the policy where students are encouraged to switch to a different program.

The rest of the paper is structured as follows. Section 2 describes the institutional context, the data and policy issues. Section 3 discusses the theoretical model and section 4 its estimation. I discuss the estimation results and the fit of the model in section 5 and I evaluate tracking policies in section 6. Finally, I conclude in section 7.

2 Institutional background and descriptive evidence

This section describes the institutional context in Flanders (Belgium) and some descriptive evidence of the impact of high school choices and results on higher education outcomes. I make use of the LOSO dataset in which I follow a sample of 5175 students that started secondary education in 1990.[6] Students were actively followed during high school and therefore the data contains many individual characteristics, choices, performance outcomes and test scores. Afterwards, the students were invited to respond to surveys about their future educational choices and labor market outcomes which provides information about their higher education career.

2.1 Study programs and the educational system

After finishing six grades in elementary school, students enroll in secondary education (high school) in the 7th grade, usually in the calendar year they become 12 years old. Students can choose between all schools in Flanders since school choice is not geographically restricted and free school choice is law-enforced.[7] In practice, most students choose one of the closest alternatives. After high school, students can enroll in higher education.

Students in full time education choose between different high school programs, grouped into tracks that differ in their academic level.[8] The academic track has the most academically rigorous curriculum. Its aim to is to provide a general education and to prepare for higher education. The middle track prepares students for different outcomes.[9] Therefore, I follow Baert *et al.* (2015) and distinguish between a track preparing mainly for higher education

[6]The LOSO data were collected by professor Jan Van Damme (KU Leuven) and financed by the Flemish Ministry of Education and Training, on the initiative of the Flemish Minister of Education.

Note that some observations were dropped because some variables were missing or because students made choices that were not consistent with the tracking systems as explained in this paper. I also restrict attention to a sample of students that did not skip a year before entering secondary education.

[7]In case capacity constraints become binding, the law protects free school choice and prevents schools from cream skimming. If the school is capacity constrained, it must add pupils to a waiting list and if spots become available, it must respect the order of this list. (http://onderwijs.vlaanderen.be/leerlingen/tien-vragen-van-leerlingen/mag-een-school-weigeren-om-mij-in-te-schrijven)

[8]Officially the distinction between tracks exists only from the third year on. However, before this, pupils decide on elective courses that prepare for a particular track.

[9]Officially this consists of two tracks called the "technical track" and "arts track".

programs (middle-theoretical), and a track that prepares more for the labor market (middle-practical). Students can also choose for the vocational track, that prepares them for specific occupations that do not require a higher education degree. Within each track, students can choose several programs which consist of bundles of elective courses. Since there are many programs to choose from, I aggregate them up to 8 study programs. In particular, I split up the academic track in four programs: classical languages, intensive math, intensive math + classical languages and other. The middle-theoretical track is split between intensive math and other.[10] This aggregation still allows for a sufficient number of students in each group and corresponds to important differences in enrollment and success rates in higher education (Declercq and Verboven, 2015).[11]

A student graduates from high school after a successful year in the 12th grade in the academic or one of middle tracks, or the 13th grade in the vocational track. Compulsory education laws require a student to pursue education until June 30th of the year he reaches the age of 18. From the age of 15, he can also decide to leave full time education and start a part-time program in which he can combine working and schooling.[12]

Despite of the fact that each track prepares for different options after secondary education, enrollment in almost any higher education option is free of selection by track or by the higher education institutes themselves. Students from any track can enroll in almost any program of higher education (Declercq and Verboven, 2018). Therefore, selection into higher education only takes the form of self-selection.

Table 1 and Table 2 summarize some stylized facts in the data on the Flemish educational system.

Table 1 looks at the distribution of students over different study programs in high school. The dataset contains measures on ability at the start of secondary education in the form of item response theory (IRT) scores based on standardized tests. These scores measure language (Dutch) and math ability and are standardized to be mean 0 and standard deviation 1. To capture differences in preferences for each program, I also include gender and socioeconomic status (SES) in the analysis. SES is measured by a dummy equal to one if at least one of the parents has completed higher education. Not surprisingly, students with high SES end up much more in the academic track. Also initial ability matters. The average language ability of a student in the academic track is 0.71 standard deviations higher than the overall

[10]On average, students follow five hours of math/week in math-intensive programs and three hours in the other programs that prepare for higher education.

[11]The supply of programs differs between schools in Flanders. Some schools specialize and offer programs in only one track while other schools do not specialize and offer programs in all tracks. In the model I will not distinguish between different schools as they are all regulated in the same way and the restrictions implied by certificates also hold for other schools.

[12]The age requirement is 16 if the student did not finish the first two grades of high school. Since success in these grades is not required, 15 years applies to the large majority of students and I will use this in the model.

Table 1: High school program and student background

Study program	Students		Male	Language ability	Math ability	High SES
All	*5175*	*(100.0%)*	*0.50*	*0.00*	*0.00*	*0.28*
Academic	*1981*	*(38.3%)*	*0.41*	*0.71*	*0.64*	*0.49*
clas+math	263	(5.1%)	0.46	1.13	1.05	0.63
clas	317	(6.1%)	0.37	0.94	0.69	0.58
math	685	(13.2%)	0.48	0.75	0.75	0.51
other	716	(13.8%)	0.32	0.41	0.36	0.38
Middle-Theoretical	*819*	(15.8%)	*0.53*	*0.11*	*0.19*	*0.22*
math	125	(2.4%)	0.70	0.32	0.47	0.30
other	694	(13.4%)	0.50	0.07	0.14	0.21
Middle-Practical	*613*	(11.8%)	*0.51*	*-0.06*	*-0.02*	*0.22*
Vocational	*1005*	*(19.4%)*	*0.51*	*-0.76*	*-0.75*	*0.10*
13th grade	610	(11.8%)	0.49	-0.67	-0.69	0.12
12th grade	395	(7.6%)	0.54	-0.89	-0.85	0.08
Dropout	*757*	*(14.6%)*	*0.67*	*-0.92*	*-0.86*	*0.07*
Part-time	434	(8.4%)	0.71	-0.96	-0.89	0.07
Full time	323	(6.2%)	0.62	-0.86	-0.81	0.08

Note: Ability measured using IRT score on tests at start of secondary education. Score normalized to be mean zero and standard deviation 1. High SES= at least one parent has higher education degree. Clas= classical languages included. Math= intensive math. Students in vocational track only obtain full high school degree after an additional 13th grade. Drop out split between students directly opting for full time dropout or first choosing part-time option.

average. For students in the vocational track this is 0.76 below the overall average. Math ability shows a similar trend.[13] Moreover, we also see large differences within tracks with both classical languages and math programs attracting stronger students. Finally, gender is important too. Male students are less likely to be in tracks that prepare better for higher education, except for math-intensive programs.[14]

Table 2: High school program and long run outcomes: summary statistics

	Higher education	
Study program	Enrollment	Degree
All	*58.1*	*44.0*
Academic	*96.9*	*84.3*
clas+math	99.2	94.3
clas	99.4	90.5
math	97.8	88.2
other	94.1	74.2
Middle-Theoretical	*82.1*	*51.8*
math	99.2	72.8
other	78.9	48.0
Middle-Practical	*54.6*	*27.4*
Vocational (13th grade)	13.4	*2.6*
Drop out	*0*	*0*

Note: Percentage of all students (also drop outs), conditional on high school program. Clas= classical languages included. Math= intensive math. Students in vocational track only obtain full high school degree after an additional 13th grade.

Table 2 looks at higher education outcomes, conditional on each high school program.

[13]For comparison, in Germany (grade 9) average language and math ability is 0.8 standard deviations above the average in the high track and 0.9 standard deviations below the average in the vocational track (Dustmann et al., 2017).

[14]The data also contains information on the location of students and schools. I use this to calculate distance to higher education options and travel time to the closest high school that offers each program.

When students leave high school, 58% decides to go directly to higher education.[15] There are large differences in higher education outcomes. 97% of all students graduating from the academic track start higher education, and the large majority eventually obtains a degree.[16] For the other tracks this is less common. Nevertheless, the track distinction is not always clear from the results as students with extra math in the middle-theoretical track obtain similar outcomes as students in the academic track without extra math or classical languages. Despite the fact that the middle-practical track prepares primarily for the job market, we do see half of the students enrolling in higher education and 27% obtains a degree.

Differences within the academic track become more clear when we look at different higher education choices. To capture heterogeneity in returns to high school study programs, I distinguish between different levels of higher education and different majors (see Appendix Table A7). Similar to Declercq and Verboven (2015), I distinguish between three types of higher education, increasing in their academic level (or prestige): professional college, academic college and university. In addition, I also distinguish between two majors in each level: STEM and other.[17]. Most students graduate from professional colleges and in a STEM major. Only students who graduated from the academic track with classical languages in their curriculum are more likely to graduate from universities than from professional colleges. Extra math in the program is associated with graduating from STEM majors and academic colleges.

2.2 Performance and the tracking policy

At the start of secondary education, all programs are available. The choice set in the future depends on the current program and performance during the year. Upward mobility, i.e. moving from a track of lower academic level to a more rigorous one, is practically impossible, (except for switches between middle tracks and the academic track in the first two grades). Similarly, students can never enroll in programs with classical languages anymore if they did

[15] To test the representativeness of the data, I compared these numbers to population data. For Belgium as a whole, I find an almost identical number of higher education enrollment around the same time period: 56% in 1996 and 57% in 1999 (UNESCO Institute for Statistics, indicator SE.TER.ENRR).

[16] As in Declercq and Verboven (2018), I define a degree as three successful years of higher education in a time span of six years.

[17] The distinction between different levels is also used in official statistics on Belgian education. To define STEM majors, I use the characterization by the Flemish government (https://www.onderwijskiezer.be/). The different types of (higher) education are also associated with large differences in wages. For descriptive purposes, I use data of the "Vacature Salarisenquete", a large survey of workers in Flanders in 2006, to compare differences in median wages. I compare the wages of 30-39 year olds (sample size of 20534 workers) with different degrees. High school drop outs earned a gross monthly wage of 2039 EUR, high school graduates without a higher education degree earned 2250 EUR, professional college graduates 2600 EUR, academic college graduates 3281 EUR and university graduates 3490 EUR. Students that graduated in a STEM major earned 3351 EUR, while students that graduated in a non-STEM major earned 2800 EUR.

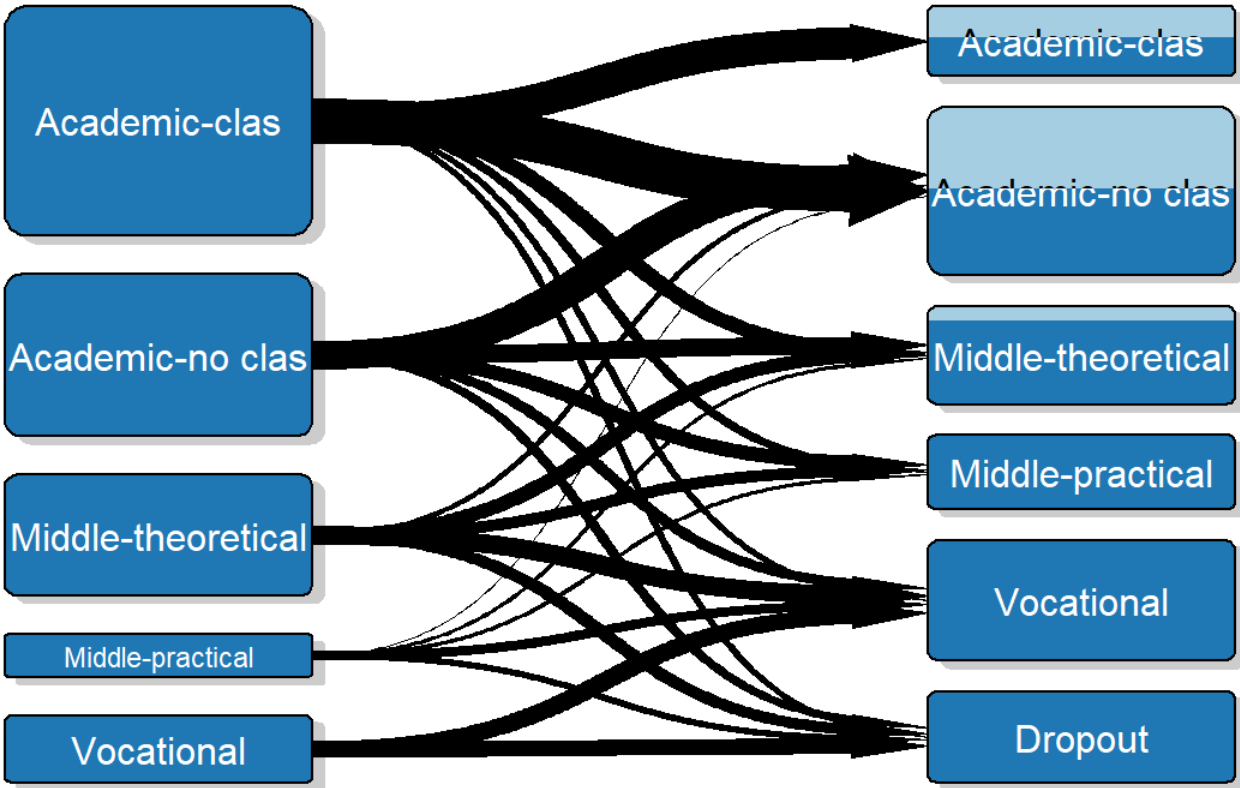
not choose it from the start. Math-intensive programs are available from grade 9 on. From then on, switching from a program without extra math to one with intensive math in the academic track is not possible. Similarly, students in the middle track that did not have extra hours of math cannot choose this anymore. Finally, there can be no more switching between full-time programs from grade 11 on.[18]

The restrictions that are imposed on student's choice sets, and the different curriculum they have from the 7th grade on, makes it important to study their decisions. A wrong choice at an early age can have large consequences for the future. As there can be uncertainty about performance and future preferences, many students like to keep their options open by choosing the academic track with classical languages in the beginning and gradually move towards their final program. Figure 1 summarizes these movements.

These transitions are not always a smooth or voluntary process. Each study program comes with it's own performance standards. Teachers are expected to teach students to help them continue in the program, but at the same time they need to keep the standards at a certain level. This is done by handing out certificates to students, based on their performance on tests during the year. An A-certificate means the student succeeded on all courses. He can then move on to the next grade and continue in the program. If he did not, teachers decide on the certificate he gets. This can still be an A-certificate, e.g. if the student only failed on a small number of courses, but it can also be a B- or a C-certificate. A C-certificate means that the student failed on too many important courses and must repeat the grade to continue in full time secondary education. A B-certificate indicates that the student failed on some important courses within the program. He can proceed to the next grade, but he will be excluded from some programs, specified in the certificate. Alternatively, a student with a B-certificate can decide to repeat the grade without being excluded from a program. In most cases, a B-certificate excludes the track a student is currently in and therefore encourages them to downgrade to another track. However, a B-certificate can also exclude certain elective courses within a track (see appendix Table A8). Most of the time students obtain an A-certificate. However, 7.1% of the certificates are B-certificates and 6.6% are C-certificates. A C-certificate always leads to grade retention if they do not want to leave full time education, but also 1 out of 4 students with a B-certificate chooses to repeat grades. Although the number of B- and C-certificates is low on a yearly basis, many students obtain at least one of them during their high school career, resulting in a large degree of grade retention. Table 3 summarizes the number of students that obtain a B- or C-certificate and

[18] Note that these rules are not always formal and students have the legal right to ignore them. Nevertheless, this is a realistic description of the perceived rules by students as schools often advertise them as being binding. Baert *et al.* (2015) apply a similar set of (informal) rules in their model.

Figure 1: Transitions from first study program to last choice in secondary education



Note: Left: program chosen in grade 7, right: last choice before leaving secondary education. Clas= classical languages included. Light blue area = proportion of students in program with extra math. See appendix Table A9 for data on these transitions.

that accumulate study delay, and compares their educational attainment with that of the average student.

We see that 32% of students leave high school with at least one year of study delay. Students with grade retention are 21 %points less likely to enroll in and 25 %points less likely to graduate from higher education than students that were not retained. Part of this is also explained by the higher drop out rates in high school. Since the negative impact of grade retention can be a causal effect, it is important that counterfactual simulations that change grade retention rates take this into account.

Table 3: Impact performance during secondary education

	Students	High school	Higher education	
		Drop out	Enrollment	Degree
All	100	14.6	58.1	44.0
At least 1 B-certificate	35.4	21.0	36.5	19.9
At least 1 C-certificate	30.0	38.6	28.9	14.7
At least 1 year of study delay	31.7	26.8	37.6	18.9

Note: Differences in outcomes for students with bad performance evaluations. A-certificate: proceed to next grade, C-certificate: repeat grade, B-certificate: repeat or downgrade.

2.3 Conclusions from descriptive evidence and implications for policy (simulations)

We observe large differences between enrollment in, but especially graduation from, higher education when we compare different study programs. More academic programs seem to encourage students to enroll in college and also help them succeed. A policy that encourages students to choose higher levels of academic effort in high school by choosing a more rigorous program is therefore expected to have a positive impact on higher education outcomes. However, at the same time we observe many students failing courses and being forced to either switch out of these programs or accumulate study delay. The latter is associated with worse outcomes. Even if these effects are causal, it is still not clear if encouraging students to choose academically rigorous programs will have an impact on higher education outcomes as for many of them it might be too costly to exert enough effort to succeed, leading to study delay or even drop out. The model that will be used for counterfactual simulations that change students incentives to exert effort should therefore not only control for the standard selection on unobservables issue, but should also identify how each policy affects academic effort through two choices the students make: the academic level of their study program, and the study effort that is exerted during the year.

I will evaluate two policies that change the choice set of students that obtain a B-certificate. In a first counterfactual, I investigate the impact of allowing students to avoid grade retention by switching to another program by simulating what would happen if a B-certificate was equivalent to a C-certificate and forced students to repeat a grade. A second counterfactual looks at an actual policy that will be implemented in Flanders and simulates the impact of not allowing students with a B-certificate to repeat the grade. This policy follows from a concern that grade retention is too high and costly for society. However, a concern with this policy is that it might reduce higher education graduation which in general is beneficial for society.

The institutional context makes it possible to do some clear counterfactuals to investigate the trade-off between costs of academic effort (study effort, grade retention, risk of drop out) and the benefits in the long run (higher education degree). Nevertheless, similar issues arise in other countries that track students from an early age like Germany, Austria or the Netherlands. In more comprehensive educational systems like the US, we find a similar trade-off at the course-level. Students often retake failed classes to graduate from high school or to increase their chances to be admitted to college. Many colleges explicitly ask for a high GPA and a rigorous academic curriculum in their admission criteria. Students, especially those of lower ability, then face a similar trade-off between studying advanced courses at the risk of retakes and a lower GPA, or choosing a curriculum with less advanced courses.

Although similar issues arise in other educational systems, they are particularly important in the current context. Belgium spends 2.8% of its GDP on secondary education, the highest number among OECD countries. It is therefore crucial to study the effectiveness of the system in helping students to achieve their future goals in a cost-efficient way. Since 96% of the cost is paid by society, it is also important to see if students have the right incentives within the system to optimize total welfare (OECD, 2017). As they do not pay the cost of schooling, they are likely to accumulate too much study delay by not exerting enough study effort during the year or choosing study programs in which they do not have a high chance to succeed. Finally, Belgium has a very high rate of grade retention in secondary education which comes at a large cost. The total cost of a year of study delay in Belgium amounts to at least 48 918 USD (corrected for PPP)/student or 11% of total expenditures on compulsory education, the highest rate in the OECD (OECD, 2013).

3 Dynamic model of academic effort choice

This section introduces a dynamic model of educational choices in which students make yearly decisions on their amount of academic effort. They make a discrete choice by choosing among a set mutually exclusive study programs and they make a decision on a continuous variable that will be an index that influences the end-of year performance of the student. I explicitly model these two yearly decisions using a structural model, but only for the time students are attending high school and their decision in the period they leave high school. As I explain in this section, this allows me to be more flexible on assumptions regarding wage expectations and higher education success and is sufficient to perform counterfactual simulations that will only affect the post-high school careers through their impact on high school study program and accumulated study delay.

Throughout the model, i refers to a student, t the time period in years, $j = 1, ..., J$ are mutually exclusive study programs in secondary education or higher education and $j = 0$ an outside option, i.e. not attending school. After each year t in high school, students obtain a performance outcome g_{it+1} that will define their choice set Φ_{it+1}. I first explain the process of performance and define a "variable effort component" that provides a tool for students to shift the performance distribution. I then discuss the utility of a study program that includes a fixed effort cost and a variable cost that increases in expected performance. I specify the value of leaving secondary education by looking at college enrollment decisions and I explain how the model is solved and how fixed and variable costs can be identified.

Details that are specific for the application like the choice set and functional form assumptions are left out of the main text but are discussed in the appendix section A.1.

3.1 End-of-year performance

Before discussing the program choice and the associated costs, I first discuss the process of performance, conditional on the study program that is chosen. Define a variable effort component $y_{it} \in (0, +\infty)$ to be scalar, continuous choice variable that influences the distribution of a discrete end-of-year performance measure $g_{it+1}.g_{it+1} = \{1, 2, 3, 4, 5\}$ defines the tracks that are available in the next grade and can be created using the data on choices in t and the certificates students obtain at the end of the year. The lowest outcome (1) does not allow any track in the next grade, and each increase corresponds to a track of higher academic level being available (vocational (2), middle-practical (3), middle-theoretical (4) and academic (5)). I abstract here from the fact that B-certificates can also exclude elective courses instead of tracks but I add this to the model as explained in appendix section A.1.

Let performance measure g_{it+1} in $t + 1$, after choosing program j in t, be the result of y_{it} and an iid shock η_{it+1} such that:

$$g_{it+1} = \bar{g} \text{ if } \bar{\eta}_j^{\bar{g}} < \ln y_{it} + \eta_{it+1} \leq \bar{\eta}_j^{\bar{g}+1} \tag{1}$$

where $\bar{\eta}_j^{\bar{g}}$ denotes the program-specific threshold to obtain at least outcome $\bar{g}$.[19] Note the difference in timing between the choice variable y_{it} and the shock η_{it+1}. Students choose y_{it} , jointly with the program j, at time t. However, they do not know the realization of g_{it+1} because of the shock η_{it+1}. This shock captures uncertainty in grading standards

[19] In the model I also allow these thresholds to vary by grade but I abstract from this in the exposition of the model for clarity of notation.

or unexpected events during the year. The information students do have at time t is the probability of obtaining an outcome $\bar{g}$ if they choose y_{it} in program j:

$$P_j(g_{it+1} = \bar{g}|y_{it}) = F(\ln y_{it} - \bar{\eta}_j^{\bar{g}}) - F(\ln y_{it} - \bar{\eta}_j^{\bar{g}+1})$$

with $F(.)$ the cumulative distribution function of the distribution of the shock. Setting $\bar{\eta}_j^1 = -\infty$ and $\bar{\eta}_j^{G+1} = +\infty$ guarantees that all probabilities add up to 1. I assume η_{it+1} is logistically distributed such that the probability of each outcome has a simple closed form solution:

$$P_j(g_{it+1} = \bar{g}|y_{it}) = \frac{\exp(\ln y_{it} - \bar{\eta}_j^{\bar{g}})}{1 + \exp(\ln y_{it} - \bar{\eta}_j^{\bar{g}})} - \frac{\exp(\ln y_{it} - \bar{\eta}_j^{\bar{g}+1})}{1 + \exp(\ln y_{ij} - \bar{\eta}_j^{\bar{g}+1})}. \tag{2}$$

Performance is then the result of an ordered logit model with index $\ln y_{it}$. Since (1) remains equivalent when adding or subtracting the same term on all sides, I can normalize one of the thresholds $\bar{\eta}_j^2 = 0$ such that all other thresholds should be interpreted with respect to the threshold of obtaining at least $g_{it+1} = 2$. Rewriting (2) for the lowest outcome ($g_{it+1} = 1$) shows that the variable effort component y can be interpreted as the odd of avoiding the lowest outcome (i.e. obtaining a C-certificate: no track in the next grade is available):

$$y_{it} = \frac{1 - P_j(g_{it+1} = 1|y_{it})}{P_j(g_{it+1} = 1|y_{it})}. \tag{3}$$

By choosing the odd of avoiding the lowest outcome, students can change the probability of each performance outcome. If y is close to zero, they are very likely to obtain the worse outcome. If y is large, they will probably reach the best outcome. It is important to note that we have not yet specified the cost of setting different levels of y. Two students with the same y have the same distribution of performance, regardless of their ability. y should therefore not be confused with other notions of effort like total hours of studying, as this is expected to have a different effect on performance, depending on the student's ability. It is rather a tool for agents in the model to choose the distribution of performance. In the next part of this section it will be clear that ability will matter for the performance distribution, not through a direct effect, but through the cost of setting the level of y.

3.2 Current and future value of a study program

Each year, students solve a dynamic problem.[20] They choose the study program j with the highest expected lifetime utility, in which the variable effort component y is chosen such that

[20]The decision will often be a collective decision by parents and their child, after advice from teachers. I do not distinguish between these different actors and simply assume some utility function is optimized, regardless of who makes the decision. See Giustinelli (2016) for a paper that does makes this distinction.

it maximizes the value of the program. The value of each program can be represented by a Bellman equation:

$$\begin{aligned} &v_{ijt}(x_{it}, \nu_i, y_{it}) + \varepsilon_{ijt} \\ &= u_j(x_{it}, \nu_i, y_{it}) + \beta \sum_{\bar{g}} P_j(g_{it+1} = \bar{g}|y_{it}) \bar{V}_{t+1}(x_{it+1}(\bar{g}), \nu_i) + \varepsilon_{ijt} \text{ for } j \in se \end{aligned} \tag{4}$$

with $v_{ijt}(x_{it}, \nu_i, y_{it})$ the conditional value function for a student i with observed state variable x and unobserved type ν of choosing program j and variable effort y at time t and $\bar{V}_{t+1}(x_{it+1}, \nu_i) \equiv \int V_{t+1}(x_{it+1}, \nu_i, \varepsilon_{it+1}) h(\varepsilon_{it+1}) d\varepsilon_{it+1}$. Students do not know future realizations of taste shocks but they do know the distribution $h(\varepsilon_{ijt})$. I follow Rust (1987) and assume this is iid extreme value type 1 distributed. The observed state variable contains the entire information set students and the econometrician share. This includes the observed student background but also time-varying and endogenous variables like past choices and performance. Because all shocks in the model are assumed to be iid, the unobserved type will capture persistent differences between students that are not captured by the observables (e.g. unobserved ability).

The conditional value function is decomposed into the flow utility of schooling, $u_j(x_{it}, \nu_i, y_{it})$, and the discounted expected value of behaving optimally from $t+1$ on, with $\beta \in (0, 1)$ the one-year discount factor.[21] The value of behaving optimally in the future is given by $V_{t+1}(x_{it+1}, \nu_i, \varepsilon_{it+1})$. This includes study costs in future years in education, but also expected outcomes after high school like the value of college enrollment, wages and leisure in the future. The expected value of the future can be written as a weighted sum over the ex-ante value functions $\bar{V}_{t+1}(x_{i,t+1}, \nu_i)$, i.e. the value functions integrated over the iid shocks in state x_{it+1}. Since performance measure g is the only stochastic element in x, the weights are simply the ordered logit probabilities of the performance measure $P_j(g_{it+1}|y_{it})$.

3.3 The utility function

Similar to Keane and Wolpin (1997), I interpret the negative of the flow utility as an effort cost of going to school. The authors then propose a functional form for this that remains constant in counterfactual simulations. Also in papers where performance plays an important role, researchers have followed this approach, while simultaneously estimating an exogenous law of motion on performance.[22] Since there is no link between costs of schooling and the

[21] I follow Arcidiacono *et al.* (2016) and set the discount factor $\beta = 0.9$.

[22] Examples of performance in these models are grade equations in Eckstein and Wolpin (1999) and Arcidiacono (2004), course credit accumulation in Joensen and Mattana (2017) and Declercq and Verboven (2018), college admission probabilities in Arcidiacono (2005) or length of study in Beffy *et al.* (2012).

performance distribution, this excludes any effect of counterfactual simulations through a direct effect on the performance distribution, e.g. because of a change in unobserved study effort.

In order to endogenize the distribution of performance, I instead propose a different form for the utility function that explicitly takes into account the student's variable effort component, i.e. the scalar they can choose to influence their distribution of performance. I split up the costs of schooling into a fixed cost $C_j^0(x_{it}, \nu_i)$, and a variable cost $c_j(x_{it}, \nu_i)y_{it}$ of effort:

$$u_j(x_{it}, \nu_i, y_{it}) = -C_j^0(x_{it}, \nu_i) - c_j(x_{it}, \nu_i)y_{it}. \tag{5}$$

The fixed cost includes cost components that are not associated with performance. This captures a (dis)taste to go to school, which is expected to depend on student and family characteristics because of differences in preferences or social norms, but also on the travel time to school. This cost can also be negative because students might enjoy going to school or parents can reward (or force) them to go. $c_j(x_{it}, \nu_i)$ is the marginal cost of effort, or the cost of increasing the variable effort component y_{it} by 1 unit. The marginal cost captures that, conditional on future values, students dislike the effort that is required to perform better.[23] Here we mainly expect differences between high and low ability students as it will be easier to obtain better performance outcomes if a student is of high ability.[24]

Note that this functional form implies a constant marginal cost assumption. However, this does not imply that there is a constant cost to increase the probability to perform better. As discussed in the previous subsection, the variable effort component is defined as the odd of avoiding the lowest performance outcome (see (3)). If the probability of the lowest outcome is already low, it is much more costly to further decrease it, than if the probability is high.[25] The model is therefore consistent with the notion that increasing the probability of being successful in the program becomes more costly when this probability is already high. At the same time, the constant marginal cost assumption will have important benefits for the estimation of the model.

[23] Similar to fixed costs, marginal cost should be interpreted as net effects. If parents encourage their children to study hard or some children enjoy it more than others, marginal costs will be lower. In contrast to fixed costs, marginal costs do need to be positive to ensure no one is willing to exert infinite effort.

[24] Dubois *et al.* (2012) also allow performance to be endogenous in their theoretical schooling model. They do not estimate the parameters of the model but show that performance in school depends on dynamic incentives created by the Progressa cash transfer program in Mexico. There is also a related literature on structural models of job search. Some models endogenize the probability to find a job, or search intensity, by estimating a cost to improve their chances (Paserman (2008), van den Berg *et al.* (2015) and Cockx *et al.* (2018)). This paper applies a similar identification strategy for marginal costs, but within a rich model of educational decisions. The richness of the model comes at the cost of computational complexity. In section 4 I describe how this burden can be reduced significantly by avoiding to solve the model during estimation.

[25] E.g. the cost of decreasing the probability to avoid the lowest outcome from 50% to 40% is 10 times smaller than decreasing it from 20% to 10%.

3.4 The value of leaving secondary education, college enrollment and college graduation

I assume leaving secondary education is a terminal action, i.e. a student never returns to high school. Students either leave the education system or (if they obtained a high school degree) they choose one of the college options. If they enter college, we could specify a utility function that looks similar to the one in high school. If they do not enter college, there is some probability to earn a wage if they find a job. In order to avoid making assumptions on how students expect their wages and college success to evolve, I directly parameterize the expected lifetime utility instead of deriving it from a model. In particular I assume the conditional value functions take the following form in $t = T_i^{SE} + 1$, with T_i^{SE} the last period student i spends in high school:

$$v_{ijt}(x_{it}, \nu_i, y_{it}) = \text{Degree}'_{it}\mu^{\text{degree}} + \Psi_j^{HEE}(x_{it}, \nu_i) \text{ if } t = T_i^{SE} + 1 \tag{6}$$

with $\text{Degree}'_{it} \subset x_{it}$ is a vector of dummy variables for each the different high school degrees a student can obtain, μ^{degree} a vector of parameters to estimate and $\Psi_j^{HEE}(.)$ a reduced form function of the state variables that predicts the higher education enrollment (HEE) decision.[26] If the student obtained a high school degree, j is a specific college option (major, level and campus), or an outside option of which the utility, net of the value of a degree, is normalized:$\Psi_0^{HEE} = 0$. If the student did not obtain a high school degree, he can only obtain the value of $j = 0$. By normalizing $\Psi_0^{HEE} = 0$, we should interpret all cost parameters in high school as the one period change with respect to the lifetime value of leaving high school without a degree. Note that this includes potential wages. A high cost of schooling can therefore also be interpreted as an opportunity cost.

Is is easy to identify $\Psi_j^{HEE}(.)$ from choices made by students that obtained a high school degree. However, this is not the case for the common parameter μ^{degree} as only differences in utility are identified. Nevertheless, we can identify it using the dynamics of the model. Students who are more likely to graduate because of higher expected performance, or because they have almost completed all grades of high school, will be less likely to drop out. This is similar to Eckstein and Wolpin (1999) who use data in secondary education to identify the value of graduating from high school. Also an exclusion restriction in the higher education choices: distance to college, will help for identification as it will encourage students differently to obtain a degree.

[26]In the model I estimate a common value of a high school degree, an interaction effect with the academic level of the program and I estimate separate effects for finishing 12th grade in the vocational track and obtaining a high school degree in the vocational track due to the specific nature of the vocational track that requires students to study an additional year in order to obtain the degree.

Since we are also interested in graduation rates from higher education, I simultaneously estimate the parameters of a reduced form model $\Psi_j^{HED}(x_{it}, \nu_i)$ that predicts graduation in each campus-level-major combination, conditional on students characteristics, high school program and study delay, and on the enrollment decision. The details are explained in the appendix section A.1.

3.5 Solving the model

It is now possible to solve the model by backward induction. I assume it is no longer possible to go to secondary education in $T^{\max} = 10$. Because of the extreme value assumption on ε_{it}, I can write the expected value of lifetime utility in the period where secondary education is no longer allowed using the logsum formula:

$$\bar{V}_t(x_{it}, \nu_i) = \gamma + \ln \sum_{j \in \Phi(x_{it})} \exp(\text{Degree}'_{it}\mu^{\text{degree}} + \Psi_j^{HEE}(x_{it}, \nu_i)) \text{ if } t = T^{\max} + 1$$

with $\gamma \approx 0.577$ the Euler constant and $\Phi_{it} = \Phi(x_{it})$ the choice set. $\bar{V}_t$ is used as an input in $t-1$ (see (4)). First, students look for the optimal value of their academic effort, conditional on each program choice. This is equivalent to finding the optimal value of the variable effort component in every possible option in secondary education: y^*_{ijt-1}. This implies the following first-order condition:

$$c_j(x_{it-1}, \nu_i) = \beta \sum_{\bar{g}} \frac{\partial P_j(g_{it} = \bar{g}|y_{it-1})}{\partial y_{it-1}} \bar{V}_t(x_{it}(\bar{g}), \nu_i) \text{ if } y_{it-1} = y^*_{ijt-1}. \tag{7}$$

The optimality condition is an equalization of marginal costs and (expected) marginal benefits. Note that we implicitly assumed the variable effort component is only affecting end-of-year performance but not the value of behaving optimally in the future, conditional on performance. Therefore we only need to take into account the derivative of the performance measure and not how future values would react directly to a change in y. In appendix section A.1.5 I show that the marginal benefits are always positive and decreasing in y. They follow an S-shaped curve, bounded by 0 and a weighted sum of the gains of obtaining a better performance measure.

For a binary performance measure, optimal effort in each program has a simple analytic solution for y, with an intuitive interpretation:

$$y^*_{ijt-1} = \sqrt{\frac{\beta(\bar{V}_t(x_{it}(2), \nu_i) - \bar{V}_t(x_{it}(1), \nu_i))}{c_j(x_{it-1})}} - 1.$$

y in period $t-1$ increases in the discounted benefits of obtaining performance level 2 instead of 1 in period t, and decreases in marginal costs c in period $t-1$. This shows a clear dynamic trade-off. Extra effort at time $t-1$ is costly but generates benefits in t. Because $y \in (0, +\infty)$, an interior solution is required. We see that this assumption puts an upper bound on marginal costs. If marginal costs are larger than the discounted benefit of having a better outcome, the student would have no incentive to exert effort. A similar intuition applies when the performance outcome is discrete (see appendix section A.1.5).[27]

When students know the optimal levels of effort in each program, they can choose the program with the highest value of $v_{ijt-1}(x_{it-1}, \nu_i, y^*_{ijt-1}) + \varepsilon_{ijt-1}$. This results in the following logit choice probabilities:

$$\Pr(d^j_{it-1} = 1 | x_{it-1}, \nu_i) = \frac{\exp(v_{ijt-1}(x_{it-1}, \nu_i, y^*_{ijt-1}))}{\sum_{j' \in \Phi(x_{it-1})} \exp(v_{ij't-1}(x_{it-1}, \nu_i, y^*_{ij't-1}))}$$

with v_{ijt-1} given by (4) for $j \in se$ and (6) for $j \in \{0, he\}$. $\bar{V}_{t-1}$ can also be calculated using:

$$\bar{V}_{t-1}(x_{it-1}, \nu_i) = \gamma + \ln \sum_{j \in \Phi(x_{it-1})} \exp(v_{ijt-1}(x_{it-1}, \nu_i, y^*_{ijt-1})).$$

These steps can be repeated until the start of secondary education to solve the entire model.

3.6 Identification of different cost component and the unobserved types

I first discuss identification if the type of a student, ν_i, is observed by the econometrician and then discuss how types can be identified if they remain unobserved. As is common for dynamic discrete choice models, identification of flow utility parameters depends on the distributional assumptions on iid shocks ε_{ijt}, only differences in utility are identified and require the normalization of the flow utility of one option ($j = 0$) (Magnac and Thesmar, 2002). The discount factor β is set before estimation.[28] The added complexity in this paper, is that I split up the flow utility of a study program in two components that depend on the state variables: a fixed cost $C^0_j(x_{it}, \nu_i)$ and a variable cost as a function of marginal costs $c_j(x_{it}, \nu_i)$. Nevertheless, I do not estimate more parameters than a model that does not make this distinction. This is because other models estimate the law of motion of performance measures directly, instead of assuming it is generated by the structural model. Instead of using data on program choices and measures of performance to estimate flow utility and a law of motion, I use it to estimate a component that is independent of performance: fixed costs,

[27] A sufficient behavioral condition that is standard in dynamic discrete choice models and implies this assumption is to assume that a student always believes there is a non-zero probability of avoiding the worst performance outcome.

[28] The discount factor can be identified using exclusion restrictions. However, the same variation is already used to identify the value of a degree (Eckstein and Wolpin, 1999).

and a component that becomes more costly when the probability to perform better increases. Optimal behavior implies that the marginal costs of the latter term can be identified from marginal benefits that arise naturally in a dynamic model (see equation (7)). This avoids the need for exclusion restrictions.

The following example shows how data on program choices and data on performance outcomes identify different parameters. Take students whose characteristics are different but they predict the same choices and outcomes from $t + 1$ on for each realization of the performance outcome g_{it+1}. This implies that they have the same marginal benefits (as a function ofy_{ijt}) at time t. If we now observe a different distribution of performance g_{it+1} after choosing the same study program at time t, i.e. they choose different optimal levels of effort in the same program, it can only be explained by differences in marginal costs. After accounting for differences in the future values, the distribution of performance and the marginal costs, differences in fixed cost $C_j^0(x_{it}, \nu_i)$ are still needed to rationalize different program choices at time t.

I now turn to the identification of the unobserved types. Note that all shocks in the model are assumed to be iid. This holds for the flow utility shocks ε_{ijt}, performance shocks η_{it} and shocks on obtaining a higher education degree. Any correlation we see in the data that cannot be captured by observable characteristics, will therefore help in identifying the unobserved type. For example, when we observe two students with identical observable characteristics, but one consistently outperforms the other during high school, it reveals something about the student's unobserved ability that will help him in high school, but might also help him in higher education. A second source of identification are exclusion restrictions in the model. In particular, I will assume that travel time to high school options influences selection into programs but has no direct effect on outcomes after secondary education.[29] This instrumental variable strategy helps in separately identifying the unobserved types from the effect of a high school program. If students living nearby specific programs obtain better outcomes, the model must attribute this effect to the program and not to unobserved ability because unobserved ability is assumed to be uncorrelated with distance and time, while they do influence program choices.

[29]See De Groote and Declercq (2018) for a discussion on the validity of distance to school as an instrument for school choice in this context. This paper assumes that distance to school is uncorrelated with success in high school. Here I make a weaker assumption by saying it is uncorrelated with success after high school. I also make an additional assumption, saying that there is no direct effect of distance on choices after high school, after controlling for distance to higher education.

4 Estimation

This model can be matched to the data, using maximum likelihood. This estimates all parameters of the model, after imposing functional form assumptions on fixed costs $C_j^0(.)$, marginal costs $c_j(.)$ and the reduced form functions for higher education enrollment and graduation. At the same time, a distribution of types ν can be estimated. However, solving the model and estimating the parameters at the same time is computationally costly, especially with unobserved types. Even in models without the additional choice of the performance distribution, many dynamic models on educational decisions have relied on methods that avoid solving the model when estimating parameters (see e.g. Arcidiacono *et al.* (2016), Declercq and Verboven (2018) and Joensen and Mattana (2017)). With the additional choice of the variable effort component, the computational issues increase further because state transitions that follow from performance can no longer be estimated in a first stage as performance is an endogenous function of the optimal choice of academic effort. Nevertheless, I can show that solutions to avoid fully solving the model, pioneered by Hotz and Miller (1993) and extended to allow for persistent unobserved heterogeneity ν_i in Arcidiacono and Miller (2011), can still be applied.

I explain first how to estimate the model if the econometrician observes the type ν of each student and then generalize to the case where ν is unobserved.

4.1 Estimation when student type is observed

To match the model to the data, the following program choice probabilities are used in the likelihood function:

$$\Pr(d_{it}^j = 1 | x_{it}, \nu_i) = \frac{\exp(v_{ijt}(x_{it}, \nu_i, y_{ijt}^*))}{\sum_{j' \in \Phi(x_{it})} \exp(v_{ij't}(x_{it}, \nu_i, y_{ij't}^*))}. \tag{8}$$

For $j \in \{0, he\}$ we can substitute in (6), which is a function of observed variables and parameters to estimate. For $j \in se$ this is not possible as (4) does not only depend on a function of parameters to estimate, but also on optimal behavior in the future and the optimal level of effort, conditional on the program choice. We would therefore have to solve the model as explained in section 3.5 to be able to estimate all parameters. To reduce the computational burden, I avoid solving the model during estimation. I first show how the first-order condition (7) allows me to substitute out the marginal cost function, when estimating other parameters of the model and how the optimal level of effort within each program can be derived directly from the data. I can then apply the CCP method of Hotz and Miller (1993) to also avoid solving for the optimal program choices during estimation.

Finally, I show how additivity of the likelihood function allows for each equation in the model to be estimated separately.

Optimal effort in each program

The first-order condition (7) helps estimation for two reasons. First, it shows that students with the same state vector (x_{it}, ν_i) at time t will choose the same effort levels in a given program: $y^*_{ijt} = y^*_{jt}(x_{it}, \nu_i)$, allowing us to recover it directly from the data. This follows from the assumption that marginal costs do not contain an unobserved shock and all future unobserved shocks are uncorrelated with the current shock. If we substitute $y^*_{jt}(x_{it}, \nu_i)$ in (3), we obtain:

$$y^*_{jt}(x_{it}, \nu_i) = \frac{1 - P_j(g_{it+1} = 1 | x_{it}, \nu_i)}{P_j(g_{it+1} = 1 | x_{it}, \nu_i)}. \tag{9}$$

If ν_i is observed, $y^*_{jt}(.)$ is easily obtained from the probability to obtain the lowest performance outcome for each realization of the state variables in the data in each period. After obtaining $y^*_{jt}(.)$, the probabilities to reach other performance levels can be used to recover the thresholds $\bar{\eta}_j = \{\bar{\eta}^1_j, ..., \bar{\eta}^G_j\}$.

A second reason the first-order condition (7) is helpful is that it allows us to write the conditional value functions (4) without a marginal cost function to estimate:

$$\begin{aligned} &v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) + \varepsilon_{ijt} \\ &= -C^0_j(x_{it}, \nu_i) \\ &+ \beta \sum_{\bar{g}} \left[\bar{V}_{t+1}(x_{it+1}(\bar{g}), \nu_i) \left(P_j(g_{it+1} = \bar{g} | y^*_{jt}(x_{it}, \nu_i)) - \frac{\partial P_j(g_{it+1} = \bar{g} | y_{ijt})}{\partial y_{ijt}} |_{y_{ijt} = y^*_{ijt}} y^*_{jt}(x_{it}, \nu_i) \right) \right] + \varepsilon_{ijt} \end{aligned} \tag{10}$$

with $y^*_{jt}(x_{it}, \nu_i)$ given by (9). $\frac{\partial P_j(g_{it} = \bar{g} | y_{ijt})}{\partial y_{ijt}}$ can be derived from the distributional assumptions on the performance measure. The conditional value function is now written with the same unknowns as in standard dynamic models with exogenous state transitions, following Rust (1987). The only difference is the transition matrix. This matrix characterizes how current states impact utility in the future. In a model with exogenous performance, this depends only on how states transition in the data. Since the distribution of performance is now chosen through y, it now also depends on y and how it affects state transitions.

CCP method

Hotz and Miller (1993) introduced the CCP (Conditional Choice Probability) method as an alternative to solve dynamic discrete choice models during estimation. They show that the future value term can be written as follows:

$$\bar{V}_{t+1}(x_{it+1}, \nu_i) = \gamma + v_{id^*t+1}(x_{it+1}, \nu_i) - \ln \Pr(d^*_{it+1} | x_{it+1}, \nu_i) \tag{11}$$

with d^*_{it+1} the vector of dummy variables for each option in which the indicator of one arbitrary option is set to 1 and $v_{id^*t+1}(.)$ the conditional value function of this option. This is particularly useful when there is an option that terminates the (structural) model. If it is possible to leave secondary education in $t+1$, we can choose $j=0$ as the arbitrary choice and substitute its value function (6) in (11):

$$\overline{V}_{t+1}(x_{it+1},\nu_i) = \gamma + \text{Degree}'_{it}\mu^{\text{degree}} + \Psi_0^{HEE}(x_{it+1},\nu_i) - \ln \Pr(d^0_{it+1}=1|x_{it+1},\nu_i). \tag{12}$$

As explained in section 3.4, $\Psi_0^{HEE}(.)$ is normalized to zero. We can now substitute (12) in (10), such that for all $j \in se$:

$$\begin{aligned} &v_{ijt}(x_{it},\nu_i,y^*_{ijt}) + \varepsilon_{ijt} \\ &= -C^0_j(x_{it},\nu_i) + \beta\gamma \\ &+ \beta \sum_{\bar{g}} \left[\begin{array}{c} \left(\text{Degree}'_{it}(\bar{g})\mu^{\text{degree}} - \ln \Pr(d^0_{it+1}=1|x_{it+1}(\bar{g}),\nu_i)\right) \\ \left(P_j(g_{it+1}=\bar{g}|y^*_{jt}(x_{it},\nu_i)) - \frac{\partial P_j(g_{it+1}=\bar{g}|y_{ijt})}{\partial y_{ijt}}|_{y_{ijt}=y^*_{ijt}} y^*_{jt}(x_{it},\nu_i)\right) \end{array} \right] + \varepsilon_{ijt}. \end{aligned} \tag{13}$$

The benefit of using the outside option $j=0$ as the arbitrary choice is that this removes the future value terms in the current period conditional value functions. This is because the terminal nature of $j=0$ allows us to write its conditional value function directly as a function of observables and parameters (see section 3.4). As in Hotz and Miller (1993), a nonparametric estimate of the Conditional Choice Probability (CCP) $\Pr(d^0_{it+1}=1|x_{it+1},\nu_i)$ can be recovered from the data, before estimating the model.[30] Because of compulsory education laws, the outside option is not always in the choice set. In the appendix, I show how the concept of finite dependence, introduced in Arcidiacono and Miller (2011), can be used to overcome this problem without fully solving or simulating the model during estimation.[31]

Likelihood function

We can now use (13) to estimate the fixed cost parameters of $C^0_j(.)$, a functional form for higher education enrollment and degree $\Psi_j^{HEE}(.), \Psi_j^{HED}(.)$ and the common component of the value of a high school degree μ^{degree}. This can be done by using the probabilities according to the model and data on program choices and long run outcomes. Let the fixed cost parameters in $C^0_j(.)$ and the common value of a degree μ^{degree} be given by μ and the reduced form parameters by ϕ. Assuming iid observations, the loglikelihood of the data is:

$$\ln L(\mu,\phi) = \sum_{i=1}^{N} \left(\sum_{t=1}^{T_i^{SE}} \ln L_{it}^{program}(\mu,\phi) + \sum_{w\in W} \ln L_i^w(\phi^w) \right)$$

[30]Similar to Arcidiacono *et al.* (2016), I estimate a flexible conditional logit to obtain predictions of the CCPs.

[31]See also Arcidiacono and Ellickson (2011) for an overview on the benefits of using finite dependence.

with $L_{it}^{program}(\mu, \phi)$ given by logit choice probabilities (8) and $L_i^w(\phi)$ given by the assumed processes on long run outcomes (see appendix section A.1). This loglikelihood could be maximized directly to obtain the estimates of (μ, ϕ). However, because of additive separability, consistent estimation could also be performed in sequential steps by first estimating the process of each long run outcome and then estimate the structural model. After estimation, the marginal cost function $c_j(.)$ can be recovered from the first-order condition (7) without requiring any additional structure on its functional form. That being said, some other parametric assumptions are often needed in applied work. While both the CCP term and the optimal level of y could be recovered nonparametrically, in this application there are not enough observations for each unique realization of x_{it} to obtain a reliable estimate of them. Therefore, flexible functional forms are imposed to predict them from the data. Similar to Arcidiacono *et al.* (2016) I will use a flexible logit to approximate the CCPs. To recover the optimal levels of y, I will use the log of a flexible index of an estimated ordered logit model on performance outcomes. This also estimates the thresholds for obtaining the different performance outcomes.

4.2 Estimation when student type is unobserved

To allow for persistent unobserved heterogeneity, I follow Arcidiacono and Miller (2011) and estimate a finite mixture of types. I assume there are $M = 2$ unobserved types m in the population, with an estimated probability to occur π_m. For interpretability, I model the types as independent from observed student background. A dummy for belonging to type 2 then enters each part of the model as if it were an observed student characteristic. To avoid an initial conditions problem, I condition the type distribution on the age the student starts secondary education: age_start_i. This is because students who accumulated study delay before secondary education will be faced with different opportunities in the model because they will be able to drop out more quickly. Since starting age depends on past grade retention, it is likely correlated with unobserved ability, creating a bias in the estimates. By conditioning the unobserved types on age_start_i, we can allow for this correlation.[32] The loglikelihood function then becomes:

$$\ln L(\phi, \mu) = \sum_{i=1}^{N} \left(\ln \sum_{m=1}^{M} \pi_{m|age\_start} \prod_{t=1}^{T_i^{SE}} L_{it}^{program,m}(\mu, \phi) L_{it+1}^{performance,m} L_{it+1}^{ccp,m} \prod_{w \in W} L_i^{w,m}(\phi^w) \right).$$

There are three main changes to the likelihood function. First, $\pi_{m|age\_start}$ is added as additional parameters to estimate, and likelihood contributions are conditioned on the

[32]This is similar to Keane and Wolpin (1997), who start their model at age 16 and condition the types on the educational attainment at that age.

type. Second, the likelihood contribution of the performance outcome in secondary education $L_{it+1}^{\text{performance},m}$ and the CCP predictor $L_{it+1}^{ccp,m}$ is added to the likelihood function. This is needed to recover the optimal levels of y in the data and the CCPs as they depend on the unobserved type. Therefore they can no longer be recovered from the data before estimating other parameters. Third, the function is no longer additively separable such that sequential estimation is not possible anymore.

Additive separability can be restored using the estimator of Arcidiacono and Miller (2011). The estimation procedure is an adaptation of the EM algorithm. The algorithm starts from a random probability of each observation to belong to each type. The entire model can then be estimated as explained in section 4.1, but weighs each observation-type combination by the probability that the observation belongs to the type. Afterwards, the joint likelihood of the data conditional on each type, is used to update the individual type probabilities, conditional on the data, using Bayes rule. This is repeated until convergence of the likelihood function. I use the two-stage estimator of Arcidiacono and Miller (2011) which implies that in the calculation of the joint likelihood, reduced form estimates of the CCPs are used for $L_{it}^{program,m}$, instead of the choice probabilities from the structural model. This means that only the population type probabilities $\pi_{m|age\_start}$, the reduced form parameters ϕ, the CCPs, the optimal effort levels in each program in the data $y_{jt}^*(x_{it}, \nu_i)$ and the thresholds $\bar{\eta}_j$ are identified in a first stage.[33] In a second stage, the fixed cost parameters and the common component of the value of a degree μ can be recovered using the structural model. Finally, the first-order condition (7) is used to recover the marginal costs. Standard errors are be obtained by using a bootstrap procedure.[34]

5 Estimation results

I present the structural schooling cost estimates, the estimates of long run outcomes and conclude with a model validation exercise.

5.1 Cost of schooling

Table 4 and Table 5 summarize the main cost estimates. While the functional form

[33] Note that the parameters of long run outcomes are already identified from the first stage, without specifying the economic structure of the model in secondary education. As mentioned by Arellano and Bonhomme (2017), this is a specific case of a nonlinear panel data model where structural assumptions are not needed to recover the parameters of interest. Therefore, it is robust to model assumptions about forward looking behavior, or rational decision making. We do however need this structure to recover the deep parameters that govern the costs of schooling. Also for counterfactual analyses, the full model is needed.

[34] I sample students with replacement from the observed distribution of the data. Since the EM algorithm takes a long time to converge, I do not correct for estimation error in the probabilities to belong to each type.

Table 4: Costs of schooling: student characteristics

	Fixed costs		Log of marginal costs	
	coef	se	coef	se
Male	11.904	(53.975)	0.306	(0.305)
x grade	-7.183	(9.871)	0.104*	(0.060)
x level	17.869**	(6.969)	0.155*	(0.093)
x math	-80.895***	(25.308)	0.474	(0.332)
x clas	-10.907	(14.484)	0.664	(0.516)
Language ability	24.021	(36.128)	-0.986***	(0.182)
x grade	-3.613	(7.448)	0.102**	(0.041)
x level	-40.234***	(8.435)	-0.130	(0.119)
x math	-2.658	(27.929)	-0.771*	(0.460)
x clas	-74.533***	(17.393)	-0.847*	(0.482)
Math ability	19.936	(33.977)	-0.423**	(0.173)
x grade	-3.255	(6.844)	0.045	(0.038)
x level	-28.703***	(7.335)	-0.355***	(0.067)
x math	-160.067***	(32.574)	0.441	(0.449)
x clas	-28.053*	(15.051)	-0.772	(0.572)
High SES	-94.599	(108.035)	-0.377	(0.377)
x grade	10.799	(20.475)	-0.044	(0.065)
x level	-26.187***	(9.706)	-0.041	(0.124)
x math	-66.789***	(18.047)	0.218	(0.329)
x clas	-44.225**	(17.750)	-0.674	(0.470)
Type 2	-413.325***	(98.484)	-2.288***	(0.451)
x grade	74.520***	(19.096)	-0.015	(0.099)
x level	-73.355***	(13.035)	-0.058	(0.130)
x math	-208.282***	(33.421)	-0.097	(0.405)
x clas	-85.433***	(19.019)	-0.594	(0.486)
Time	1	(.)	-0.001	(0.001)

Note: Estimates of a sample of 5175 students or 33357 student-year observations during secondary education. Fixed cost estimates of equation (15), divided by μ_{time} such that they can be interpreted in minutes of daily travel time. . The marginal costs in the model are a nonparametric function of state variables, this table summarizes them by regressing its logarithmic transformation on the same variables that enter the fixed costs. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Grade subtracted by 6 to start counting in secondary education. Clas= classical languages included. Math= intensive math. Level = academic level of high

Table 5: Costs of schooling: tracking

	Fixed costs		**Log of marginal costs**	
	coef	se	coef	se
Repeat grade	305.477***	(86.990)	-1.433***	(0.498)
x grade	16.942	(13.750)	0.431***	(0.112)
x level	117.605***	(25.817)	-0.451***	(0.165)
Study delay	93.395*	(49.519)	0.970***	(0.302)
x grade	-17.217*	(9.185)	-0.073	(0.060)
x level	2.170	(7.947)	0.025	(0.080)
Downgrade	229.078***	(40.222)	0.179	(0.122)
Upgrade	449.730***	(67.277)	0.386	(0.246)

Note: Estimates of a sample of 5175 students or 33357 student-year observations during secondary education. Fixed cost estimates of equation (15), divided by μ_{time} such that they can be interpreted in minutes of daily travel time. . The marginal costs in the model are a nonparametric function of state variables, this table summarizes them by regressing its logarithmic transformation on the same variables that enter the fixed costs. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Grade subtracted by 6 to start counting in secondary education. Clas= classical languages included. Math= intensive math. Level = academic level of high school program (0-3). Common components of costs are reported in appendix Table A2. Bootstrap standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

assumption on the fixed costs is the same as shown in these tables, the marginal costs are derived from optimal choices in the data and are therefore a nonlinear function of the variables and other parameters in the model (see section 4). I therefore perform an OLS regression on the logarithmic transformation of the estimated marginal costs with the same structure as the fixed costs to interpret them. The fixed cost parameters are divided by the parameter of the travel time variable and can therefore be interpreted in terms of daily minutes to travel.[35]

Table 4 shows that the effect of student characteristics on fixed and marginal costs. Some characteristics mainly affect fixed costs. We see that male students have a lower fixed costs for math programs, equivalent to a decrease in daily travel time of 81 minutes, while their effect on other marginal and fixed costs are smaller and more imprecisely estimated. High SES studens are more favorable towards programs of higher academic level, more math and with classical languages. While gender and SES seem to be important contributors to the fixed costs, their effects on marginal costs are smaller compared to other characteristics, and often not significant. Our measures of initial ability have a strong impact on marginal costs. An increase by 1% of a standard deviation in language ability decreases marginal costs by 0.9% in grade 7 and still about 0.3% towards the end of high school. The same increase in math ability decreases marginal costs by about 0.4% over the entire high school career of students in the vocational track (level =0), but by almost 1.2% in the academic track (level=3). Finally, we see that there is still a lot of persistent heterogeneity in the data that our observable characteristics are not capturing. Appendix Table A1 shows that 67% of students belong to type 1 and 33% to type 2. Type 2 students have much lower fixed costs of schooling, especially at the start of high school and in rigorous programs. Moreover, the marginal costs of a type 2 students are only 10% (=exp(-2.288)) of those of a type 1 student.

Table 5 shows the impact of track choices and grade retention during high school. We see that study delay, i.e. past grade retention, increases marginal costs. This could be the result of demotivation, the same increases in expected performance might be perceived more costly for a student with study delay because he might loose interest in studying for tests. On the contrary, we find very strong decreases in marginal costs when students are repeating a grade, especially at the start of high school and in programs of high academic level. This can results from the fact that students see the same course material for a second time, making it easier to succeed. While it might be easier to succeed, many students dislike repeating a grade as is demonstrated by the large increase in fixed costs. This shows a clear trade-off: students dislike repeating a grade, but it does help them to perform well, especially in more academic programs. This is one of the explanations why students might consider repeating a

[35] Choice-specific constants and grade interaction can be found in the appendix Table A2.

grade, even if they have the possibility to go to the next grade in another program. Finally, students do not like to switch programs. Both down- and upgrading is associated with much higher fixed costs, indicating a preference of students to stay in the same program.[36]

5.2 Long run outcomes

All estimates for long run outcomes can be found in appendix section A.3. Important here is that student characteristics that were important in explaining the costs of schooling also have a direct effect on college enrollment and graduation. This is also the case for type 2 students which makes it important to control for unobserved heterogeneity when assessing the causal impact of study programs. Because the estimates of high school programs on long run outcomes are difficult to interpret, I also calculate the total Average Treatment Effects on the Treated (ATT) of each study program (see Table 6) as follows:

$$ATT^{j'} = E_{x,\nu}\left[P_j^{HE}(x_{it_{HE}}(j'), \nu_i) - P_j^{HE}(x_{it_{HE}}(j^0), \nu_i) | d^{j'}_{iT_i^{SE}} = 1\right] \text{ for } HE = \{HEE, HED\} \quad (14)$$

with $E_{x,\nu}$ an expectations operator over the empirical distribution of the observables x and the estimated distribution of the unobserved types ν. P_j^{HE} is the probability of the enrollment decision or higher education degree outcome of each college option as a function of the state variables. $x_{it_{HE}}(j')$ is the observed state vector of student i in the data at the time the outcome is realized $t = t_{HE}$ and $x_{it_{HE}}(j^0)$ is the same vector but with the graduation track replaced by an arbitrary benchmark program j^0. The ATT then calculates the average effect on HE of graduating from j' instead of j^0 for the group of students who graduated from j' in the data. The estimate is not the results of a counterfactual simulation but it is a "ceteris paribus" causal effect, i.e. it is the effect of one variable if all other variables that were realized before leaving secondary education are kept fixed. Similarly, I calculate the effect of one year of study delay by comparing outcomes for retained students in the counterfactual scenario where they would not have accumulated study delay.

Most estimates point in the same direction as a simple comparison of means in the data, but to a smaller extent. I find that graduating from the academic track without classical languages or extra math leads to an increase in college graduation of 26 %points compared to the middle-practical track. Also the other higher education oriented track, the middle-theoretical track, leads to lower chances of college graduation (14 %points). Elective courses mainly matter for type of higher education but there we also see large differences in graduation rates among graduates from both the academic and middle-theoretical track that had classical languages or intensive math in their program.

[36] Note that the cost of upgrading is only identified in the first two grades because it is not allowed afterwards.

Table 6: ATTs of high school program and delay in higher education (in %points difference)

	Higher education			
	Enrollment (% of students)		Degree (% of students)	
	coef	se	coef	se
Study program				
Academic				
clas+math	1.1***	(0.2)	5.4***	(2.2)
clas	1.4***	(0.3)	5.9***	(2.3)
math	1.4***	(0.2)	5.3***	(1.8)
other	*benchmark*		*benchmark*	
Middle-Theoretical				
math	0.2	(0.112)	-3.9	(4.3)
other	-7.9***	(2.186)	-13.7***	(2.7)
Middle-Practical	-28.6***	(2.992)	-25.6***	(3.1)
Vocational	-66.3***	(3.696)	-44.3***	(3.7)
One year of study delay	0.8	(2.281)	-9.3***	(1.9)
Data	58.1		44.0	

Note: Average treatment effects on the treated (ATT). Clas= classical languages included. Math= intensive math. Effects on enrollment, first-year success and degree completion after graduating from different high school programs, compared to graduating from the academic track without clas or math option, and the effects of one year of study delay, compared to 0. Effects are calculated using indexes, specified in appendix section A.1, for each individual at the realization of other variables. Effects on obtaining higher education degree take into account effects through enrollment. Bootstrap standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Tracking policies alter the trade-off between the academic level of the program a student is in and years of study delay he accumulates. It is therefore also interesting to look at the ATT of study delay. I compare the effect for students with zero and one year of study delay for those that are retained during secondary education. I do not find a significant effect on higher education enrollment but there is a strong negative impact on obtaining a higher education degree. One year of study delay decreases the probability to obtain a higher education degree by 9 %points.

5.3 Model fit

After estimation, I solve the model as explained in section 3.4. Once I have solved the model backwards to find all the conditional value functions and variable effort levels, I forward simulate all error terms to simulate choices of study program and the distribution of performance. I then compare this database to the actual data and use it to compare it to counterfactual scenarios. To allow students to change their effort level in a given program, I perform a grid search over different levels of the variable effort component in each conditional value function to look for the optimal value.[37] Table 7 shows the ability of the model to replicate the actual data. The model does a decent job in predicting the patterns in the data such that we can use it for counterfactual simulations. Nevertheless, there are some small differences between the data and the model. There is a slight overprediction of graduating from the academic and middle-theoretical track, and an underprediction in the middle-practical track. There is also a small overprediction in the number of students that obtain study delay, despite the smaller prediction errors in obtaining certificates that could lead to grade retention (B and C-certificates). Higher education outcomes are predicted very precisely.

[37] I predict value functions for a sample of students and weigh them according to the empirical distribution of the discrete variables in the data, and a distretized transformation of the continuous variables. Within each group, I take draws of the continuous variables and ensure that no draw ever represents more than 50 students. To forward simulate error terms, I replicate each draw by the number of students it represents to obtain a database of the same size as the original sample. This procedure has the benefit of having sufficient draws, while needing only a limited number of students to use for a grid search of the optimal effort level in each program. The grid search for effort levels starts at the optimal level in the data and looks for better levels with increments in the log of effort of 0.05 with a minimum of -5 and a maximum of +5.

Table 7: Predictions of the model

	Data	Predictions	
High school (% of students)			
Academic	38.3	*40.4*	*(2.3)*
clas+math	5.1	3.3	(0.6)
clas	6.1	3.2	(0.5)
math	13.2	15.7	(0.9)
other	13.8	18.3	(1.7)
Middle-Theoretical	15.8	17.0	(1.3)
math	*2.4*	*2.8*	(0.5)
other	13.4	14.2	(1.0)
Middle-Practical	11.8	7.8	(1.4)
Vocational	19.4	20.8	(1.0)
Dropout	*14.6*	*13.9*	(0.7)
Students with at least 1 B-certificate	35.4	36.8	(0.9)
Students with at least 1 C-certificate	30.0	*30.9*	*(1.2)*
Students with at least 1 year of study delay	31.7	34.4	(1.4)
Higher education (% of students)			
Enrollment	58.1	*59.4*	*(1.0)*
Graduation	44.0	44.9	(1.0)
University degree	12.4	11.4	(0.6)
Academic college degree	6.1	6.2	(0.4)
Professional college degree	25.5	27.3	(0.9)
Degree in STEM major	16.9	16.9	(0.7)

Note: Clas= classical languages included. Math= intensive math. Observed outcomes in the data and prediction from a dynamic model with program and effort choice in secondary education. High school graduation summarizes the programs in which students graduated or dropout and the number of students with grade retention. Bootstrap standard errors in parentheses.

6 Counterfactual tracking policies

In the current tracking policy in Flanders, teachers decide if a student has acquired the necessary skills to transition to the next grade in each of the programs. In some cases, students have not acquired the skills to transition to the next grade, regardless of their program choice. They then obtain a C-certificate which requires them to repeat the grade. However, in many cases students are allowed to transition to the next grade but have to switch to a program of lower academic level.[38] In this case they obtain a B-certificate. This allows underperforming students to avoid grade retention, however they can still opt for the same program if they are willing to repeat the grade. In a first counterfactual, I look at the effect of this policy by removing the option to avoid grade retention and force students to repeat the grade if they underperformed during this year. In a second counterfactual, I instead reinforce this policy of avoiding grade retention, by forcing them to downgrade instead of having a choice between repeating the grade or downgrading.

I first discuss the predicted effect of each policy and then show the biases in a model where performance is assumed to be exogenous.

6.1 Changes in the B-certificate policy

In Table 8 I compare the outcomes of the two counterfactuals to the status quo for the aggregate statistics and for the group of students that obtain a B-certificate in the status quo scenario. Appendix Table A10 shows how the number of students who graduate in each program changes. The policy "Repeat" forces students to repeat the grade after obtaining a B-certificate, i.e. removing downgrading as a way to avoid grade retention. The policy "Downgrade" forces students to downgrade after a B-certificate by not allowing them to repeat the grade.

Without the ability to avoid grade retention, outcomes would have been worse. The policy does not manage to increase graduation rates from the academic track, instead we see an increase in drop out by 3.3 %points. Not surprisingly, the number of students with grade retention increases by a large amount (8.1 %points). It is therefore not surprising that higher education outcomes are worse. Enrollment decreases by 1.7 %points and graduation by 1.8 %points. I conclude that the current policy is therefore better than a strict pass or fail policy. Nevertheless, it can be further improved.

If students who obtained a B-certificate were not allowed to repeat the grade, grade retention would decrease by 10.7 %points and drop out by 1.4 %points. This does come

[38] In some cases, they only need to drop an elective course.

Table 8: Predictions of the model: counterfactual tracking policy

	Status quo	Policy change B-certificate			
		Repeat		Downgrade	
	(%)	(Change in %points)		(Change in %points)	
Panel A: All students					
High school					
At least 1 B-certificate	36.8	-10.7***	(0.6)	-3.8***	(0.5)
At least 1 C-certificate	30.9	-0.2	(0.2)	-2.2***	(0.3)
At least 1 year of study delay	34.4	8.1***	(0.5)	-10.7***	(0.7)
Dropout	13.9	3.3***	(0.3)	-1.4***	(0.3)
Higher education					
Enrollment	59.4	-1.7***	(0.3)	-1.7***	(0.3)
Graduation	44.9	-1.8***	(0.3)	-0.3	(0.3)
Panel B: Students with B-certificate in status quo					
High school					
At least 1 B-certificate	100.0	-31.0***	(1.7)	-12.1***	(1.2)
At least 1 C-certificate	38.0	0.7	(0.6)	-4.4***	(0.7)
At least 1 year of study delay	54.3	22.8***	(1.2)	-27.2***	(1.4)
Dropout	20.2	9.0***	(0.8)	-3.6***	(0.6)
Higher education					
Enrollment	42.6	-3.0***	(0.6)	-3.7***	(0.7)
Graduation	26.3	-3.6***	(0.7)	-0.2	(0.7)

Note: Predictions from a dynamic model in which students choose study program and distribution of performance. B-certificate = students acquired skills to proceed to next grade but only in track of lower academic level or if they drop elective course. Status quo = students can choose to downgrade or repeat grade after obtaining B-certificate, Repeat = students must repeat grade after obtaining B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining B-certificate. Bootstrap standard errors in parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

at a cost in the short run. Students switch to programs of lower academic level, which decreases enrollment in higher education by 1.7 %points. However, this policy only decreases enrollment for students with low chances of eventually graduating from college such that graduation rates remain the same.

Although the new system affects every student in their decisions (as they all have some probability to obtain a B-certificate), we see that the results are largely driven by the students that currently obtain a B-certificate. Note however than many of them no longer obtain one in a counterfactual scenario. Both counterfactuals make it less favorable for students to have a B-certificate and we see that students avoid them, either by choosing a different track or by studying harder. This is especially the case in the repeat counterfactual where almost one out of three students with a B-certificate in the status quo no longer obtains one in the counterfactual. Without these behavioral effects, the impact of the counterfactual would be even worse.

I conclude that the current tracking policy is a good alternative to guide students in their track choices, rather than having them repeat a grade if they fail. Nevertheless, the choice they currently have to repeat a grade instead of downgrading does not lead to beneficial effects in the long run. Given a large cost of grade retention and drop out for society, a policy in which students are not allowed to repeat a grade after having received a B-certificate should be considered.

6.2 Bias in model with exogenous performance

Traditional dynamic models of educational decisions model performance as exogenous, i.e. conditional on the program choice and the state variables, the distribution of performance is given. Both performance and utility are then estimated as functions of the state variables and these functions are kept fixed in counterfactual simulations. This is a strong restriction on the behavior of the student. Most policies that are expected to change program choices because of dynamic considerations, are also expected to change study effort and therefore performance. The counterfactuals considered here give students an incentive to study harder as they can avoid a B-certificate that would restrict their choice set more than in the status quo. I assess the importance of this aspect in the model by comparing the estimates of my model with that of a model where the performance distribution remains the same mapping from state variables to performance as in the status quo and the marginal cost in the utility function is set to 0 such that (new) estimates of the fixed costs capture the entire utility. Table 9 summarizes the results on the main outcomes.

We see that a model without endogenous performance leads to less favorable outcomes

Table 9: Predictions of the model: bias in policy simulation with exogenous performance

	Policy change B-certificate							
	Repeat (Change in %points)				Downgrade (Change in %points)			
Endogenous performance	Yes	No	Bias		Yes	No	Bias	
Panel A: All students								
High school								
At least 1 B-certificate	-10.7	-7.3	3.3***	(0.5)	-3.8	-0.5	3.2***	(0.5)
At least 1 C-certificate	-0.2	-0.0	0.2	(0.2)	-2.2	-1.5	0.7***	(0.2)
At least 1 year of study delay	8.1	10.9	2.8***	(0.4)	-10.7	-9.3	1.4***	(0.4)
Dropout	3.3	5.1	1.8***	(0.2)	-1.4	-1.1	0.3**	(0.1)
Higher education								
Enrollment	-1.7	-2.9	-1.2***	(0.2)	-1.7	-2.6	-0.9***	(0.1)
Graduation	-1.8	-2.8	-1.0***	(0.2)	-0.3	-1.2	-0.9***	(0.1)
Panel B: Students with B-certificate in status quo								
High school								
At least 1 B-certificate	-31.0	-21.4	9.6***	(1.2)	-12.1	-1.9	10.2***	(1.2)
At least 1 C-certificate	0.7	0.9	0.2	(0.6)	-4.4	-3.7	0.7	(0.4)
At least 1 year of study delay	22.8	30.3	7.5***	(1.0)	-27.2	-25.0	2.2**	(0.9)
Dropout	9.0	14.2	5.3***	(0.5)	-3.6	-3.2	0.4	(0.3)
Higher education								
Enrollment	-3.0	-6.1	-3.2***	(0.4)	-3.7	-6.8	-3.1***	(0.4)
Graduation	-3.6	-6.2	-2.5***	(0.4)	-0.2	-3.1	-2.9***	(0.3)

Note: All counterfactuals allow for bevioral effects of students through study program choice but only with endogenous performance they also choose the distribution of performance. B-certificate = students acquired skills to proceed to next grade but only in track of lower academic level or if they drop elective course. Status quo = students can choose to downgrade or repeat grade after obtaining B-certificate, Repeat = students must repeat grade after obtaining B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining B-certificate. Bootstrap standard errors in parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

in both counterfactuals. For example, the increase in drop out rates in the repeat policy is 5 %points instead of 3, and the decrease in the downgrade policy is 1.1 %points instead of 1.4. Also in higher education we see more negative effects if performance is exogenous and we would falsely conclude that there is a negative impact on higher education graduation from the downgrade policy if we would not take endogenous performance into account.

The results can be explained by an increase in study effort of students. Both counterfactuals make it less favorable to obtain a B-certificate. In a dynamic model with program choice, students can avoid this by choosing a program in which their success rate is higher. If performance is endogenous, they could also change the success rate itself. Although it will be costly to do so, for many students this might be a better option than to switch programs or risk failure. This extra incentive to exert study effort leads to a decrease in B-certificates that is largely neglected by a model with exogenous performance. This is very clear for the students that obtained a B-certificate in the status quo. In the downgrade policy, a model with exogenous performance predicts 98% of them would still have a B-certificate, but with endogenous performance this decreases to 88%. In the repeat policy it goes from 79% to 69%.

We can conclude that it is important to allow for students to change their effort in counterfactual simulations, not only through their choice of study program, but also through their study effort during the year by endogenizing the distribution of performance.

7 Conclusion

I estimated a dynamic model of effort choice in secondary education in which students choose the academic level of the study program, as well as the distribution of their performance. Using a dataset for Flanders (Belgium), I find that academic programs are important for educational attainment. Nevertheless, policies that encourage students who underperform to opt for programs of lower academic level do not have a negative effect on obtaining a higher education degree and they significantly decrease grade retention and drop out from high school. This shows that small changes to tracking policies in secondary education can have important effects. Future research should therefore focus more on how different tracking policies can improve student outcomes.

From a methodological perspective, I show that it is important and feasible to control for the fact that students can change their effort levels within a program in response to counterfactual policy changes. Study effort allows students to avoid some of the negative consequences of imposed policies. This is especially the case when performance is expected to increase due to more strict policies, possibly through unobserved study effort. This turned out to be important in the current analysis. Further research can apply the modeling strategy to other contexts where dynamics are important. Any model where agents are expected to to have some, but imperfect, control over state transitions can benefit from this approach and the data requirements are the same as for a model with exogenous state transitions. E.g. in the seminal paper of Rust (1987), agents decide on when to replace bus engines, depending on current and expected future mileage and engine prices. Mileage is treated as stochastic, but it is reasonable to assume that its predictable component is the choice variable of a firm that wants to maximize the profits of letting the buses drive. An increase in future prices of bus engines can then have an impact on the miles buses are expected to drive, which would change the state transitions in the model.

References

Altonji, J. G. (1995). The Effects of High School Curriculum on Education and Labor Market Outcomes. *The Journal of Human Resources*, 30(3):409.

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics*, 4(1):185–223.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, 73(5):1477–1524.

Arcidiacono, P., Aucejo, E., Maurel, A., and Ransom, T. (2016). College Attrition and the Dynamics of Information Revelation. *Working paper.*

Arcidiacono, P. and Ellickson, P. B. (2011). Practical Methods for Estimation of Dynamic Discrete Choice Models. *Annual Review of Economics*, 3:363–394.

Arcidiacono, P. and Miller, R. A. (2011). Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity. *Econometrica*, 79(6):1823–1867.

Arellano, M. and Bonhomme, S. (2017). Nonlinear panel data methods for dynamic heterogeneous agent models. *Annual Review of Economics*, 9:471–496.

Aughinbaugh, A. (2012). The effects of high school math curriculum on college attendance: Evidence from the NLSY97. *Economics of Education Review*, 31(6):861–870.

Baert, S., Cockx, B., and Picchio, M. (2015). Modeling the Effects of Grade Retention in High School. *Working paper.*

Beffy, M., Fougere, D., and Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter? *Review of Economics and Statistics*, 94(1):334–347.

Cockx, B., Dejemeppe, M., Launov, A., and Van der Linden, B. (2018). Imperfect monitoring of job search: Structural estimation and policy design. *Journal of Labor Economics*, 36(1).

Costrell, R. M. (1994). A Simple Model of Educational Standards. *American Economic Review*, 84(4):956–971.

De Groote, O. and Declercq, K. (2018). Tracking and specialization of high schools: heterogeneous effects of school choice. *Working paper*.

Declercq, K. and Verboven, F. (2015). Socio-economic status and enrollment in higher education: do costs matter? *Education Economics*, 23(5):532–556.

Declercq, K. and Verboven, F. (2018). Enrollment and degree completion in higher education without admission standards. *Economics of Education Review*, 66:223–244.

Dubois, P., de Janvry, A., and Sadoulet, E. (2012). Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico. *Journal of Labor Economics*, 30(3):555–589.

Dustmann, C., Puhani, P. A., and Schoenberg, U. (2017). The Long-term Effects of Early Track Choice. *The Economic Journal*, 127(603):1348–1380.

Eckstein, Z. and Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1339.

Garibaldi, P., Giavazzi, F., Ichino, A., and Rettore, E. (2012). College cost and time to complete a degree: Evidence from tuition discontinuities. *Review of Economics and Statistics*, 94(3):699–711.

Giustinelli, P. (2016). Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of The High School Track. *International Economic Review*, 57(2):573–601.

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. A. and Woessman, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. *The Economic Journal*, 116(510):C63–C76.

Heckman, J. J., Humphries, J. E., and Veramendi, G. (2016). Dynamic treatment effects. *Journal of Econometrics*, 191(2):276–292.

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529.

Joensen, J. S. and Mattana, E. (2017). Student Aid, Academic Achievement, and Labor Market Behavior. *Working paper*.

Joensen, J. S. and Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, 44(1):171–198.

Keane, M. P. and Wolpin, K. I. (1997). The Career Decisions of Young Men. *Journal of Political Economy*, 105(3):473–522.

Kreisman, D. and Stange, K. (2017). Vocational and Career Tech Education in American High Schools: The Value of Depth Over Breadth. *Education Finance and Policy*, Forthcoming.

Magnac, T. and Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816.

Meer, J. (2007). Evidence on the returns to secondary vocational education. *Economics of Education Review*, 26:559–573.

Mehta, N. and Fu, C. (2018). Ability Tracking, School and Parental Effort, and Student Achievement: A Structural Model and Estimation. *Journal of Labor Economics*.

Nord, C., Roey, S., Perkins, R., Lyons, M., Lemanski, N., Brown, J., and Schuknecht, J. (2011). The Nation's Report Card [TM]: America's High School Graduates. Technical Report NCES 2011-462, US Department of Education, National Center for Education Statistics, Washington, DC.

OECD (2013). *PISA 2012 Results: What Makes Schools Successful (Volume IV)*. PISA. OECD Publishing.

OECD (2017). *Education at a Glance 2017*. Education at a Glance. OECD Publishing.

Paserman, M. D. (2008). Job search and hyperbolic discounting: Structural estimation and policy evaluation. *The Economic Journal*, 118(531):1418–1452.

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.

Rose, H. and Betts, J. R. (2004). The effect of high school courses on earnings. *The Review of Economics and Statistics*, 86(2):497–513.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Todd, P. E. and Wolpin, K. I. (2018). Accounting for Mathematics Performance of High School Students in Mexico: Estimating a Coordination Game in the Classroom. *Journal of Political Economy*, (forthcoming).

van den Berg, G. J. and van der Klaauw, B. (2015). Structural Empirical Evaluation of Job Search Monitoring. *Working paper.*

A Appendix

A.1 Details of of the model

This section discusses the details of the model that were left out of the paper. I first describe how the institutional context influences the choice set and I discuss functional form assumptions that are used in estimation. I then describe the shape of the marginal benefits function and explain how the model can be estimated in periods where there is no drop out option due to compulsory education.

A.1.1 Choice set

After completion of elementary education, students start in secondary education. Each year in secondary education, they choose a study program $j \in se$. Each study program belongs to one of four tracks: academic (*acad*), middle-theoretical (*midt*), middle-practical (*midp*) and vocational (*voc*). Within the academic track, students can also choose for math-intensive programs (*math*), and/or classical languages (*clas*) in the curriculum. In the middle-theoretical track they can also choose for a math-intensive program. The tracks are available throughout secondary education, i.e. grade 7 to 12 (and 13 in the vocational track). The classical languages option starts at the same time, while the math options start in grade 9. Next to the full time education system, there is also a part-time vocational option (*part*). This option is available from the moment a student is 15 years old and does not have a grade structure.

The program choices are restricted. First, students can never upgrade tracks according to the following hierarchy: $acad > midt > midp > voc > part$, with the exception of the first two grades in which mobility between *acad*, *midt* and *midp* is allowed. Second, they can stop with their specialization in extra math or classical languages, but the reverse is not possible.[39] Finally, from grade 11 on, students who want to stay in full time education must stay in the same program.

Students progress in secondary education by obtaining a certificate at the end of the year. As explained in the institutional context, the flexibility of a B-certificate can have different implications on the choice set. I therefore use the certificate data to create a variable that captures the permission for a student to enter or continue in each track in the next grade as a measure of performance g_{it+1}. To allow for B-certificates to only exclude elective courses, I extend the model to allow for additional performance measures that contain permissions to study classical languages g^{clas}_{it+1} and intensive math g^{math}_{it+1}.

[39] However it is allowed to switch from *acad* without extra math to *midt* with extra math.

From the age of 18 on, students have the possibility to leave the education system: $j \in (0, he)$ with $j = he$ enrollment in a higher education (if they obtained a high school degree) and $j = 0$ the outside option. I assume this is a terminal choice, i.e. they never return to secondary education.

A.1.2 Costs and performance

Section 3 describes the model without specifying the variables that are used in the analyses and how they impact the effort costs. The estimation section 4 explains the need for specifying a functional form for fixed costs, but marginal costs can be derived from the marginal benefits at the optimal level of effort in the data. In this section I impose functional form assumptions on fixed cost and explain the performance measure that is used to derive the optimal level of effort from the data. I also extend the model to allow for more than one measure of performance such that B-certificates can also exclude elective courses.[40]

Fixed cost of secondary education

Students pay a fixed cost for the program they choose in secondary education, unrelated to their expected performance. Note that schools in Flanders are tuition-free, so the cost is only a psychic cost. As they can also enjoy school, the sign of the fixed cost is not restricted. Let C^0_{ijt} be the fixed cost of student i in option j at time t:

$$\begin{aligned} C^0_{ijt} &= \mu^0_j + \mu^{\text{grade}}_j \text{grade}_{ijt} \\ &+ S'_i(\mu^{S,0}_j + \mu^{S,\text{grade}} \text{grade}_{ijt} + \mu^{S,\text{level}} \text{level\_SE}_{ijt} + \mu^{S,\text{math}} \text{math}_{ijt} + \mu^{S,\text{clas}} \text{clas}_{ijt}) \\ &+ \nu'_i(\mu^{\nu,0}_j + \mu^{\nu,\text{grade}} \text{grade}_{ijt} + \mu^{\nu,\text{level}} \text{level\_SE}_{ijt} + \mu^{\nu,\text{math}} \text{math}_{ijt} + \mu^{\nu,\text{clas}} \text{clas}_{ijt}) \\ &+ \mu_{time} \text{time}_{ijt} \\ &+ \text{retention}'_{ijt}(\mu^{\text{ret},0} + \mu^{\text{ret,grade}} \text{grade}_{ijt} + \mu^{\text{ret,level}} \text{level\_SE}_{ijt}) \\ &+ \mu_{up} \text{upgrade}_{ijt} + \mu_{down} \text{downgrade}_{ijt}. \end{aligned} \tag{15}$$

μ is a vector of parameters to estimate. S_i is a vector of time-invariant observed student characteristics, ν_i is a vector of dummy variables that indicate to which type the student belongs, time_{ijt} is the daily commuting time by bike to the closest school that offers the study option in the current grade and grade_{ijt} is the grade a student is in (renormalized to be 1 in the first year of high school).[41] Level_SE_{ijt} is the academic level of the track a student

[40] Note that the part-time track does not have a grade structure. I therefore only model its fixed cost. Due to a lack of variation, I only estimate a choice-specific constant, which implies that student background should have the same effect on part-time and full-time drop out.

[41] Commuting time by bike is measured by geocoding address data using the STATA command "geocode3" and by calculating travel time using the STATA command "osrmtime". A bike is the most popular mode of transporta-

is in with 0 the vocational track, 1 the middle-practical track, 2 the middle-theoretical track and 3 the academic track and math and class refer to respectively programs with intensive math and with classical languages. Study delay is captured by the 2x1 vector: retention_{ijt}. This vector contains a flow variable: a dummy equal to one if the student is currently in the same grade as the year before and a stock variable that captures the years of study delay accumulated in previous years. Finally, upgrade_{ijt} and downgrade_{ijt} are dummy variables indicating if a student is currently in a track with at a higher or lower academic level than the year before.

Note that in section 3, the scale of the utility function was implicitly normalized to unity. Therefore all parameters μ are identified. However, to directly interpret the cost estimates, I will rescale the parameters by dividing by μ_{time}. This way, the cost estimates can be measured in daily commuting time.

End-of-year performance

Study performance during the year has an impact on future utility through potential grade retention and changes in choice sets. At the end of the year, students obtain an A, B or C certificate that defines their choice set for the next grade. The main measure of performance is $g_{ijt+1} = \{1, 2, ..., 5\}$. If $g_{ijt+1}^{track} = 1$, student i's performance at time t was insufficient to go to the next grade, regardless of the program they want to follow. $g_{ijt+1}^{track} = 2$ allows access to the next grade of the vocational track (*voc*) but not other tracks. Similarly, $g_{ijt+1}^{track} = 3$ additionally allows access to the next grade *midp*, $g_{ijt+1}^{track} = 4$ allows *midt* and $g_{ijt+1}^{track} = 5$ allows *acad*. In the final year of the program, the measures no longer allow access to a certain track but result in a high school degree.

In section 4, I explained how a measure of performance can be used to back out the optimal level of the variable effort componenty_{ijt}^{*} in a nonparametric way. However, the finite number of observations and the large state space does not allow me to do this. I therefore approximate the y_{ijt}^{*} by a parametric structure.[42] The optimal levels of effort and the thresholds to obtain each outcome can then be recovered by estimating an ordered logit model with index $\ln(y_{ijt}^{*})$. Note that some of the thresholds are not identified from the data but from the institutional context that imposes restrictions on mobility. I also allow the thresholds to differ not only by different programs but also by the grade a student is

tion. According to government agency VSV, 36% of students use a bike, 30% the bus and 15% a car (source: http://www.vsv.be/sites/default/files/20120903_schoolstart_duurzaam.pdf). Since distance to school is small, travel time by bike is also a good proxy for other modes of transportation.

[42]I impose the same structure on the logarithmic transformation of effort as for the fixed costs, but I also add the effects of distance to higher education institutes and characteristics of last year's program. This is because distance should not have an effect on the fixed cost of schooling in secondary education, but it can have an effect on the optimal level of effort because future utility is affected. I add characteristics of the program a student followed in the previous year to allow experience to affect effort today because of a change in marginal cost of effort. These results can be found in the appendix section A.2.

in.[43] Note that there is still no need to impose structure on the marginal costs as they are recovered from the marginal benefits at the optimal level of y in the data (see equation (7) at page 19).

Extension to allow for course-specific restrictions

The model in section 3 only includes one measure of performance. I defined this as the permission to start in each track in the next grade. The problem with this approach is that B-certificates can also exclude elective courses instead of tracks. I therefore extend the model to allow for two additional measures of performance: $g_{it+1}^{clas} = \{1,2\}$ specifies if a students can go to the next grade in a program that includes classical languages. $g_{it+1}^{math} = \{1,2,3\}$ specifies if a student can go to the next grade in a math option in the middle-theoretical track ($g_{it+1}^{math} = 2$) or the academic track ($g_{it+1}^{math} = 3$).[44] I model their distribution by an ordered logit, conditional on the outcome of g_{it+1}, with indexes:

$$g_{it+1}^{math^\circ} + \eta_{it+1}^{math} = \alpha_y^{math} \ln y_{it} + S_i' \alpha_S^{math} + \nu_i' \alpha_\nu^{math} + \eta_{it+1}^{math} \tag{16}$$

$$g_{it+1}^{clas^\circ} + \eta_{it+1}^{clas} = \alpha_y^{clas} \ln y_{it} + S_i' \alpha_S^{clas} + \nu_i' \alpha_\nu^{clas} + \eta_{it+1}^{clas}. \tag{17}$$

I also estimate grade- and track-specific thresholds.[45] $\alpha_y^{math} > 0$ and $\alpha_y^{clas} > 0$ measure how much effort, identified from the permissions to start in each track, matters for each elective course. I also allow for comparative advantages in elective courses by estimating the influence of observed and unobserved student characteristics through $(\alpha_S^{math}, \alpha_\nu^{math})$ and $(\alpha_S^{clas}, \alpha_\nu^{clas})$.

A.1.3 Choice after leaving secondary education

From the age of 18 on, students can leave the education system. If they decide to leave, they can either go to a higher education option ($j \in he$) or choose the outside option $j = 0$ (but might enroll later). If a student leaves secondary education without a degree, he cannot go to higher education. Admission to the first year of higher education is allowed for all students with a high school degree. As explained in section 2, I distinguish between three types of higher education, increasing in their academic level: professional college, academic college

[43] Because there is little variation in the data, I do not estimate separate thresholds for each program but distinguish between thresholds in the academic track and thresholds in other tracks.

[44] The downside of this extension is that marginal benefits of effort in the model are not necessarily decreasing over the entire domain of y_{ijt}, making it more difficult to find a solution. Nevertheless, an interior solution is still required and this should satisfy the first-order condition, allowing me to estimate the model as explained in section 4, but now by calculating joint probabilities for all performance outcomes instead of one particular outcome. When solving the model in counterfactuals, I use a grid search around the optimal level in the data to find a new optimum.

[45] Note that the thresholds for elective courses are not always estimated as they can also be deterministic, given the result of g_{it+1}^{track}. If $g_{it+1}^{track} < 4$, $g_{it+1}^{math} = g_{it+1}^{clas} = 1$. If $g_{it+1}^{track} = 4$, $g_{it+1}^{math} = \{1,2\}$ and $g_{it+1}^{clas} = 1$.

and universities. In addition, I also distinguish between STEM and non-STEM majors and allow for five campuses to study a university program: Leuven, Ghent, Antwerp, Brussels and Hasselt. For professional and academic colleges, I follow Declercq and Verboven (2018) and assume students choose the closest campus. This results in 15 options after graduating from high school: $j = 0$ or one of the 14 study options $j \in he$.

As explained in section 3.4, the value functions after leaving secondary education can be written as the sum of an estimated, common value of a high school degree μ^{degree} and a choice-specific component. I now impose structure on the choice-specific component $\Psi_j^{HEE}(x_{it}, \nu_i) = \Psi_{ij}^{HEE}$ to let this value differ by the option a student chooses after graduation:

$$\begin{aligned}
\Psi_{ij}^{HEE} = &\ \phi_j^{HEE,0} \\
&+ S_i'(\phi^{HEE,S,0} + (\phi^{HEE,S,level}\text{level_HE}_j + (\phi^{HEE,S,STEM}\text{STEM}_j) \\
&+ \nu_i'(\phi^{HEE,\nu,0} + (\phi^{HEE,\nu,level}\text{level_HE}_j + (\phi^{HEE,\nu,STEM}\text{STEM}_j) \\
&+ \phi^{HEE,\text{dist}}\text{distance_HE}_{ij} \\
&+ d'_{iT_i^{SE}}\phi^{HEE,\text{SE}} + \text{delay}_{iT_i^{SE}}\phi^{HEE,delay} + \phi^{HEE,level\text{x}delay}\text{level_SE}_{iT_i^{SE}}\text{xdelay}_{iT_i^{SE}} \\
&+ X'_{ij}\phi^{HEE,\text{interact}}
\end{aligned} \tag{18}$$

Level_HE is the level of the higher education program with professional college (0), academic college (1) and university (2). Distance_HE$_{ij}$ is the distance in kilometers from the student's home to the chosen option. $d_{iT_i^{SE}}$ is a vector of dummy variables for each possible program a student can graduate in and $\text{delay}_{iT_i^{SE}}$ the years of accumulated study delay. Since there are very few students in the academic track that do not enroll in higher education, I do not distinguish between them and estimate common effect of the academic track on enrollment. I also include a vector of interactions X_{ijt} that includes all interactions between characteristics of the high school program the student graduated in (academic level, intensive math, classical languages) and the characteristics of the higher education program (level and STEM major). Since only differences in utility are identified, I normalize $\phi_0^{HEE} = 0$.

Note that travel time to secondary education programs does not enter the lifetime utility of leaving secondary education directly, while it does influence the program students choose. It therefore serves as an exclusion restriction that helps in identifying the unobserved type distribution.

A.1.4 Higher education degree and reduced form functions in counterfactual simulations

To evaluate the impact of high school programs on long run outcomes, a structural model in secondary education is needed as it allows for policy counterfactuals that will not change the primitives of the model, like the fixed cost of a study program, marginal costs of effort within a program or the value of a degree, but it will change student behavior. Without a structural model, we would not be able to assess the effects of changes in policy. For outcomes after secondary education, we do not need to know the primitives of the model but only the way these outcomes are influenced by secondary education outcomes, after controlling for observed and unobserved student characteristics. I therefore model a reduced form function only. This is similar to the approach in the dynamic treatment effect literature (Heckman et al., 2016), but I only apply it to choices after leaving high school to be able to do counterfactual simulations during secondary education in which students are forward looking.

The enrollment decision in higher education is already modeled in a reduced form way. We do not distinguish between fixed and variable costs of effort, nor do we separate flow utility for the future payoffs. The parameters of the model than simply describe how student characteristics and high school degree and study delay influence enrollment decisions. I impose a similar model for graduation from higher education. I use the same functional form as in 18 but I add interaction effects in X to take into account the enrollment decision. In particular I include dummy variables for choosing the same level, upgrading a level, and choosing the same major. I add an additional iid shock that is distributed extreme value type 1 such that I obtain logit probabilities. Since these shock are iid, it is important to take into account outcomes of the enrollment decision which can capture correlation between enrollment decision and the final degree a student obtains.

After estimation, the estimated functions of both enrollment and graduation can be used to look at the impact of counterfactual policies in secondary education. Let $x_{it_{HE}}(Policy = 0)$ be the realized state vector of i at time t_{HE} in the status quo scenario, and $x_{it_{HE}}(Policy = p')$ the state vector in the counterfactual scenario. The expected impact on the proportion of students with long run outcome HE of policy p' is then given by:

$$E_{x,\nu}\left[P_j^{HE}(x_{it_{HE}}(Policy = p'), \nu_i) - P_j^{HE}(x_{it_{HE}}(Policy = 0), \nu_i)\right] \text{ for } HE = \{HEE, HED\}$$

with $E_{x,\nu}$ an expectations operator over the empirical distribution of the observables x and the estimated distribution of the unobserved types ν. P_j^{HE} is the probability of the enrollment decision or higher education degree outcome of each college option as a function of the state variables. $x_{it_{HE}}(Policy = p')$ is the observed state vector of student i in the data at the time the outcome is realized $t = t_{HE}$ in the counterfactual policy and $x_{it_{HE}}(Policy = 0)$ is the same vector but in the status quo scenario.

A.1.5 Curvature of marginal benefits

This section discusses the shape of the marginal benefits of effort that were introduced in section 3. Note that marginal costs of effort are assumed to be constant and the variable effort component lies in the open interval $(0, +\infty)$. The marginal benefits should therefore be sufficiently flexible to guarantee an interior solution. I show that the marginal benefits are positive, decreasing in effort and follow an S-shaped curve.[46]

Marginal benefits are positive

In the paper section 3, I described that the marginal benefits of effort are given by

$$MB(x_{it}, \nu_i, y_{it}) = \beta \sum_{\bar{g}} \frac{\partial P_j(g_{it+1} = \bar{g}|y_{it})}{\partial y_{it}} \bar{V}_{t+1}(x_{it+1}(\bar{g}), \nu_i).$$

Note that $P_j(g_{it+1} = \bar{g}|y_{it}) = P_j(g_{it+1} \leq \bar{g}|y_{it}) - P_j(g_{it+1} \leq \bar{g} - 1|y_{it})$ for $g_{it+1} > 1$ and $P_j(g_{it+1} = 1|y_{it}) = P_j(g_{it+1} \leq 1|y_{it})$:

$$MB(x_{it}, \nu_i, y_{it}) = \beta \sum_{\bar{g}<G} \frac{\partial P_j(g_{it+1} \leq \bar{g}|y_{it})}{\partial y_{it}} \left(\bar{V}_{t+1}(x_{i,t+1}(\bar{g}), \nu_i) - \bar{V}_{t+1}(x_{i,t+1}(\bar{g}+1), \nu_i)\right).$$

Because performance shocks are distributed logistically, we know that $\frac{\partial P_j(g_{it+1} \leq \bar{g}|y_{it})}{\partial y_{it}} = -\frac{(1-P_j(g_{it+1} \leq \bar{g}|y_{it}))P_j(g_{it}}{y_{it}}$ Since $0 < P_j(g_{it+1} \leq \bar{g}|y_{it}) < 1$, it is sufficient to assume that students value a higher performance measure ($\bar{V}_{t+1}(x_{it+1}(g_{it+1}+1), \nu_i) > \bar{V}_{t+1}(x_{it+1}(g_{it+1}), \nu_i)$) to prove that $MB(x_{it}, \nu_i, y_{ijt}) > 0$. The last expression is also intuitive: the marginal benefit is larger with large gains of getting a higher performance outcome, but less so if effort is already high.

Marginal benefits are decreasing in effort

First note that $\beta\left(\bar{V}_{t+1}(x_{it+1}(g_{it+1}+1), \nu_i) - \bar{V}_{t+1}(x_{it+1}(g_{it+1}), \nu_i)\right)$ is always positive and does not depend on y_{it}. Therefore, a sufficient condition for the marginal benefits to be decreasing is $\frac{\partial \frac{(1-P_j(g_{it+1} \leq \bar{g}|y_{it}))P_j(g_{it+1} \leq \bar{g}|y_{it})}{y_{it}}}{\partial y_{it}} < 0 \; \forall g_{it+1} < G$:

[46] Note that in the application of the model, I extend the model to allow for additional performance measures. In rare cases this can lead the marginal benefits to be increasing in small intervals.

$$
\begin{aligned}
\frac{\partial \frac{(1-P_j(g_{it+1}\leq \bar{g}|y_{it}))P_j(g_{it+1}\leq \bar{g}|y_{it})}{y_{it}}}{\partial y_{it}} &= -\frac{\partial P_j(g_{it+1}\leq \bar{g}|y_{it})}{\partial y_{it}} P_j(g_{it+1}\leq \bar{g}|y_{it})(y_{it})^{-1} \\
&+ (1-P_j(g_{it+1}\leq \bar{g}|y_{it}))\frac{\partial P_j(g_{it+1}\leq \bar{g}|y_{it})}{\partial y_{it}}(y_{it})^{-1} \\
&- (1-P_j(g_{it+1}\leq \bar{g}|y_{it}))P_j(g_{it+1}\leq \bar{g}|y_{it})(y_{it})^{-2} \\
&= \frac{\partial P_j(g_{it+1}\leq \bar{g}|y_{it})}{\partial y_{it}}(y_{it})^{-1}(1-2\ddot{P}_j(g_{it+1}|y_{it})) \\
&- (1-P_j(g_{it+1}\leq \bar{g}|y_{it}))P_j(g_{it+1}\leq \bar{g}|y_{it})(y_{it})^{-2} \\
&= (1-P_j(g_{it+1}\leq \bar{g}|y_{it}))P_j(g_{it+1}\leq \bar{g}|y_{it})(y_{it})^{-2}(2P_j(g_{it+1}\leq \bar{g}|y_{it})-2) \\
&= -2(y_{it})^{-2}P_j(g_{it+1}\leq \bar{g}|y_{it})(1-P_j(g_{it+1}\leq \bar{g}|y_{it}))^2.
\end{aligned}
$$

Since $P_j(g_{it+1} \leq \bar{g}|y_{it}) > 0$ and $y_{it} > 0$, we find that $\frac{\partial \frac{(1-P_j(g_{it+1}\leq \bar{g}|y_{it}))P_j(g_{it+1}\leq \bar{g}|y_{it})}{y_{it}}}{\partial y_{it}} < 0$ and therefore $\frac{\partial MB(x_{it},\nu_i,y_{it})}{\partial y_{it}} < 0$, i.e. there are decreasing returns to effort.

Marginal benefits are S-shaped

Note that we can rewrite

$$
\begin{aligned}
\frac{(1-P_j(g_{it+1}\leq \bar{g}|y_{it}))}{y_{it}} &= \frac{1}{y_{it}}\left(1-\frac{\exp(\bar{\eta}_j^{\bar{g}+1}-\ln y_{it})}{1+\exp(\bar{\eta}_j^{\bar{g}+1}-\ln y_{it})}\right) \\
&= \frac{1}{y_{it}}\left(\frac{1}{1+\exp(\bar{\eta}_j^{\bar{g}+1})/y_{it})}\right) \\
&= \frac{1}{y_{it}+\exp(\bar{\eta}_j^{\bar{g}+1})}.
\end{aligned}
$$

Marginal benefits then become

$$
MB(x_{it},\nu_i,y_{it}) = \beta\sum_{\bar{g}<G}\frac{1}{y_{it}+\exp(\bar{\eta}_j^{\bar{g}+1})}P_j(g_{it+1}\leq \bar{g}|y_{it})\left(\bar{V}_{t+1}(x_{it+1}(\bar{g}+1),\nu_i)-\bar{V}_{t+1}(x_{it+1}(\bar{g}),\nu_i)\right).
$$

Because $P_j(g_{it+1} \leq \bar{g}|y_{it}) \to 1$ if $y_{it} \to 0$, the lower limit of $y_{it} \in (0, +\infty)$ is given by

$$
\lim_{y_{it}\to 0} MB(x_{it},\nu_i,y_{it}) = \beta\sum_{\bar{g}<G}\frac{1}{\exp(\bar{\eta}_j^{\bar{g}+1})}\left(\bar{V}_{t+1}(x_{it+1}(\bar{g}+1),\nu_i)-\bar{V}_{t+1}(x_{it+1}(\bar{g}),\nu_i)\right).
$$

If $y_{it} \to +\infty$, $P_j(g_{it+1} \leq \bar{g}|y_{it}) \to 0$ for $\bar{g} < G$. Therefore the upper limit is:

$$
\lim_{y_{it}\to +\infty} MB(x_{it},\nu_i,y_{it}) = 0.
$$

Because of the two asymptotes and the fact that MB are always decreasing in effort, we obtain an S-shaped curve. When effort is very high, the probability to obtain the highest performance level reaches 1, making additional effort useless. The benefit can also never be larger than the differences between the lifetime utility from obtaining a higher outcome. The larger the thresholds, the more difficult it is to obtain the higher outcome. Therefore, the differences in utility in the upper limit of marginal benefits are inversely weighted by the size of the thresholds to capture the differences in probability.

Note that these bounds are also the upper and lower limits of the marginal costs we allow for in the model since $y_{it} \in (0, +\infty)$ implies an interior solution where the marginal benefits curve crosses the constant marginal costs.

A.1.6 CCP estimation without terminal action

The CCP estimation described in the paper is only possible if students are allowed to leave secondary education in $t+1$. However, for most students we start modeling choices from the age of 12. At $t+1$, they are age 13 and do not have that option because of compulsory schooling laws. They will get the outside option $j=0$ at $t+6$. I write ρ_{it} to be the number of years it takes before the CCP correction term with the outside option can be applied: $\rho_{it} = \max\{1, 18 - Age_{it}\}$. Since ρ_{it} can be different from 1, it makes the correction term more complicated. However, the intuition is similar. We need to repeat the CCP method in future values until the outside option is available. This is an application of finite dependence, introduced in Arcidiacono and Miller (2011). In contrast to their application on problems that have a renewal action in the future, I apply it to the terminal action of choosing to work. Nevertheless, the exposition in this section is very similar to Arcidiacono and Miller (2011) and Arcidiacono and Ellickson (2011) and I refer to their papers for more details about finite dependence.

The choice probabilities (8) can also be written by using differenced value functions:

$$\Pr(d_{it}^{j} = 1 | x_{it}, \nu_i) = \frac{\exp\left(v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) - v_{ij't}(x_{it}, \nu_i, y^*_{ij't})\right)}{1 + \sum_{j^\circ \in \Phi(x_{it})} \exp\left(v_{ij^\circ t}(x_{it}, \nu_i, y^*_{ij^\circ t}) - v_{ij't}(x_{it}, \nu_i, y^*_{ij't})\right)}$$

$$\begin{aligned} &\text{with } v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) - v_{ij't}(x_{it}, \nu_i, y^*_{ij't}) \\ &= u_j(x_{it}, \nu_i) + \beta \sum_{\bar{g}} P_j(g_{it+1} = \bar{g} | y^*_{ijt}) \bar{V}_{t+1}(x_{it+1}(\bar{g})) \\ &- u_{j'}(x_{it}, \nu_i) - \beta \sum_{\bar{g}} P_{j'}(g_{it+1} = \bar{g} | y^*_{ij't}) \bar{V}_{t+1}(x_{it+1}(\bar{g})), \end{aligned} \tag{19}$$

for any $j' \in \Phi(x_{it})$ and $u_j(x_{it}, \nu_i) = -C_j^0(x_{it}, \nu_i) - c_j(x_{it}, \nu_i)y^*_{ijt}$, with $y^*_{ijt} = y^*_{jt}(x_{it}, \nu_i)$. Substitute the CCP representation of the future value as a function of the CCP of an arbitrary choice and its conditional value function (11) in (19):

$$\begin{aligned}
&v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) - v_{ij't}(x_{it}, \nu_i, y^*_{ij't}) \qquad (20)\\
&= u_j(x_{it}, \nu_i) + \beta \sum_{\bar{g}} P_j(g_{it+1} = \bar{g}|y^*_{ijt}) \left(\gamma + v_{id^*t+1}(x_{it+1}(\bar{g}), \nu_i) - \ln \Pr(d^*_{it+1}|x_{it+1}(\bar{g}), \nu_i)\right)\\
&- u_{j'}(x_{it}, \nu_i) - \beta \sum_{\bar{g}} P_{j'}(g_{it+1} = \bar{g}|y^*_{ij't}) \left(\gamma + v_{id^*t+1}(x_{it+1}(\bar{g}), \nu_i) - \ln \Pr(d^*_{it+1}|x_{it+1}(\bar{g}), \nu_i)\right)
\end{aligned}$$

with d^*_{it+1} the vector of dummy variables in which only the dummy corresponding to the arbitrary choice is equal to one, and $v_{id^*t+1}(.)$ the conditional value function of this option. Define the cumulative probability of being in a particular state given the current state variable and choice, and a particular decision sequence $d^*_i = (d_{it}, d^*_{it+1}, d^*_{it+2}, ...d^*_{it+\rho_{it}})$:

$$\begin{aligned}
\kappa^*_\tau(g_{i\tau+1}|x_{it}, \nu_i) &= \sum_{\bar{g}} P_{d^*}(g_{i\tau+1} = \bar{g}|y^*_{d^*t}(x_{i\tau}, \nu_i)) \text{ if } \tau = t \qquad (21)\\
\kappa^*_\tau(g_{i\tau+1}|x_{it}, \nu_i) &= \sum_{\bar{g}} P_{d^*}(g_{i\tau+1} = \bar{g}|y^*_{d^*t}(x_{i\tau}, \nu_i))\kappa^*_{\tau-1}(g_{i\tau}|x_{it}, \nu_i) \text{ if } \tau > t
\end{aligned}$$

with $P_{d^*}(g_{i\tau+1} = \bar{g}|y^*_{d^*t}(x_{i\tau}, \nu_i))$ the probability of receiving performance outcome $g_{i\tau+1} = \bar{g}$ at time $t = \tau$, in the program a student will be according to the decision sequence d^*_i. Similarly, define κ'_τ to be the transitions in a sequence where the choice in t is different: $d'_i = (d'_{it}, d^*_{it+1}, d^*_{it+2}, ...d^*_{it+\rho_{it}})$.[47] We can then repeat the CCP method in each of the future periods and rewrite (20) as the sum of future flow utilities and CCPs until the outside option becomes available at $t + \rho_{it}$:

$$\begin{aligned}
&v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) - v_{ij't}(x_{it}, \nu_i, y^*_{ij't}) \qquad (22)\\
&= u_j(x_{it}, \nu_i) - u_{j'}(x_{it}, \nu_i)\\
&+ \beta[u_{d^*}(x_{it+1}(g_{it+1}), \nu_i) - \ln \Pr(d^*_{it+1}|x_{it+1}(g_{it+1}), \nu_i)]\kappa^*_t(g_{it+1}|x_{it}, \nu_i)\\
&- \beta[u_{d^*}(x_{it+1}(g_{it+1}), \nu_i) - \ln \Pr(d^*_{it+1}|x_{it+1}(g_{it+1}), \nu_i)]\kappa'_t(g_{it+1}|x_{it}, \nu_i)\\
&+ \sum_{\tau=t+2}^{t+\rho_{it}-1} \beta^{\tau-t}[u_{d^*}(x_{i\tau}(g_{i\tau}), \nu_i) - \ln \Pr(d^*_{i\tau}|x_{i\tau}(g_{i\tau}), \nu_i)]\kappa^*_{\tau-1}(g_{i\tau}|x_{it}, \nu_i)\\
&- \sum_{\tau=t+2}^{t+\rho_{it}-1} \beta^{\tau-t}[u_{d^*}(x_{i\tau}(g_{i\tau}), \nu_i) - \ln \Pr(d^*_{i\tau}|x_{i\tau}(g_{i\tau}), \nu_i)]\kappa'_{\tau-1}(g_{i\tau}|x_{it}, \nu_i)\\
&+ \beta^{\rho_{it}}\overline{V}_{t+\rho_{it}}(x_{t+\rho_{it}}(g_{it+\rho_{it}}), \nu_i)\kappa^*_{t+\rho_{it}-1}(g_{it+\rho_{it}}|x_{it}, \nu_i)\\
&- \beta^{\rho_t}\overline{V}_{t+\rho_{it}}(x_{t+\rho_{it}}(g_{it+\rho_{it}}), \nu_i)\kappa'_{t+\rho_{it}-1}(g_{it+\rho_{it}}|x_{it}, \nu_i).
\end{aligned}$$

[47] We can also allow a more general alternative sequence in which the choice in each period is different but here it is sufficient to only let the first choice be different.

$\overline{V}_{t+\rho_{it}}$, the value of behaving optimally when the outside option is available, can be written as in (12). The calculation of the value function is now possible after choosing the arbitrary options in each period, the prediction of their CCPs and the predictions of optimal effort in the study program. However, further simplifications follow from a good choice of the arbitrary options and a convenient parameterization of the model.

In the paper I explained why I choose $j = 0$ when the outside option is available. The institutional context can also offer further simplifications by choosing the right programs in other periods. Since upward mobility from the lowest track is never allowed, I argue that the arbitrary choices should always be the lowest track available in each period: the vocational track if a student is not 15 years old yet, and the part-time track if the student is older. This choice significantly removes the number of CCPs and future utility terms we need. From the moment students choose the vocational track, they can no longer make choices until the part-time track becomes available. Similarly, once students opt for the part-time track, they can no longer make other choices until the outside option $j = 0$ is available. Therefore we only need a CCP at the time a student is switching tracks in the sequence. Moreover, since the part-time track does not follow a grade-structure and students can never return to the standard grade-structure, the state variables will not evolve anymore in a way that depends on choices made. Arcidiacono and Ellickson (2011) explain that in this case, the future utility terms after choosing that option can be ignored in estimation as they will cancel out in the differenced value functions.

The same procedure is applied within $u_j(x_{it}, \nu_i) = -C_j^0(x_{it}, \nu_i) - c_j(x_{it}, \nu_i)y^*_{jt}(x_{it}, \nu_i)$. By substituting the marginal cost of effort by the marginal benefit of effort, future value terms also enter directly into $u_j(x_{it}, \nu_i)$ (see (10)). Because $\sum_{\bar{g}} \frac{\partial P_j(\bar{g}|y_{ijt})}{\partial y_{ijt}} = 0$, all terms that do not depend on performance drop out such that the same simplifications arise because of finite dependence.

A.2 Additional estimation results secondary education

Table A1: Type probabilities in %

	Type probabilities	
	Type 1	Type 2
Overall	*66.53*	*33.47*
Age 12	62.64	37.36
Age 13	88.14	11.86
Age 14	88.37	11.63

Note: Estimates of unobserved types in the student population by age they start secondary education.

Table A2: Costs of schooling

	Fixed costs		Log of marginal costs	
	coef	se	coef	se
Academic				
clas+math	616.666***	(88.516)	-1.066	(1.264)
clas	196.059***	(65.352)	-0.615	(0.478)
math	316.397***	(72.299)	-0.035	(0.800)
other	-17.861	(65.612)	0.168	(0.314)
x grade	-19.348	(16.238)	0.445***	(0.088)
Middle-theoretical				
math	359.500***	(75.694)	0.973	(0.879)
other	-66.867	(67.017)	0.950***	(0.258)
x grade	-4.009	(14.603)	0.149**	(0.070)
Middle-practical	29.575	(61.419)	1.528***	(0.328)
x grade	-20.056	(12.981)	0.026	(0.084)
Vocational	32.336	(64.456)	-1.063*	(0.578)
x grade	22.992	(16.166)	0.379***	(0.115)
Part-time	321.590***	(52.933)	.***	(.)
Math x grade	-25.841***	(5.450)	-0.002	(0.142)
Clas x grade	-11.094**	(4.711)	-0.088	(0.129)

Note: Estimates of a sample of 5175 students or 33357 student-year observations during secondary education. Fixed cost estimates of equation (15), divided by μ_{time} such that they can be interpreted in minutes of daily travel time. The marginal costs in the model are a nonparametric function of state variables, this table summarizes them by regressing its logarithmic transformation on the same variables that enter the fixed costs. Grade subtracted by 6 to start counting in secondary education. Clas= classical languages included. Math= intensive math. Bootstrap standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$ (normal-based).

A.3 Estimation results long run outcomes

Table A3: Value of obtaining degree

	Degree values	
	coef	**se**
High school degree	661.447***	(177.865)
x level	110.173	(92.821)
High school degree vocational (difference)	155.339	(161.440)
12th grade certificate vocational	728.081***	(121.331)

Note: Estimates of μ^{degree} in equation (6). All parameters are divided by $-\mu_{time}$ in equation (15) such that they can be interpreted in minutes of daily travel time. Level = academic level of high school program (0-3). Bootstrap standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A4: Estimation results higher education (1)

	Higher education			
	Enrollment		Degree	
	coef	se	coef	se
Male	-1.033***	(0.153)	-0.697***	(0.141)
x HE level	0.122*	(0.062)	0.139	(0.101)
x STEM	0.964***	(0.116)	0.564***	(0.167)
Language ability	0.334**	(0.142)	0.150	(0.182)
x HE level	0.467***	(0.074)	0.309***	(0.094)
x STEM	-0.097	(0.101)	-0.121	(0.184)
Math ability	0.095	(0.109)	0.488**	(0.200)
x HE level	0.296***	(0.095)	0.153	(0.106)
x STEM	0.634***	(0.108)	-0.008	(0.212)
SES	0.559***	(0.186)	0.569***	(0.175)
x HE level	0.432***	(0.065)	0.278***	(0.089)
x STEM	0.138	(0.092)	-0.100	(0.194)
Type 2	0.474***	(0.179)	1.039***	(0.161)
x HE level	0.685***	(0.056)	0.311***	(0.092)
x STEM	1.138***	(0.094)	-0.433**	(0.171)

Note: Estimates of higher education outcomes as specified in section A.1. HE Level = level of higher education (0-2). Bootstrap standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A5: Estimation results higher education (2)

	Higher education			
	Enrollment		Degree	
	coed	se	coef	se
Academic degree	4.230***	(0.340)		
Academic + clas + math degree			-0.490	(0.517)
Academic + clas degree			0.055	(0.332)
Academic + math degree			0.176	(0.237)
Academic other degree			benchmark	
Middle-theoretical degree	3.332***	(0.258)		
Middle-theoretical + math degree	.	(.)	-0.886***	(0.326)
Middle-theoretical + other degree	.	(.)	-0.741***	(0.170)
Middle-practical degree	2.011***	(0.229)	-1.183***	(0.214)
Vocational degree	benchmark		-2.523***	(0.449)
Study delay	0.170	(0.207)	-0.477	(0.317)
High school level x study delay	-0.315***	(0.104)	-0.157	(0.142)
HE level x High school level	0.418***		-0.155	(0.107)
HE level x clas	0.869***	(0.076)	0.453***	(0.104)
HE level x math	0.382***	(0.080)	0.269***	(0.096)
HE level x study delay	-0.072	(0.072)	-0.029	(0.146)
STEM x high school level	-0.536***	(0.081)	-0.336**	(0.143)
STEM x clas	-0.419***	(0.139)	0.372	(0.256)
STEM x math	0.917***	(0.112)	0.482***	(0.175)
STEM x study dekay	-0.190	(0.116)	0.147	(0.220)

Note: Estimates of higher education outcomes as specified in section A.1. Clas= classical languages included. Math= intensive math. High school level = academic level of high school program (0-3). HE Level = level of higher education (0-2). Bootstrap standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A6: Estimation results higher education (3)

	Higher education			
	Enrollment		Degree	
	coef	se	coef	se
Distance (km)	-0.017***	(0.001)	-0.016***	(0.002)
Same HE level as enrollment			1.427***	(0.126)
Same major as enrollment			2.279***	(0.102)
Upgrade HE level			-2.013***	(0.323)
University	-8.099***	(0.422)	-4.627***	(0.691)
Academic college	-5.035***	(0.251)	-3.367***	(0.399)
Professional college	-1.825***	(0.214)	-2.213***	(0.222)
STEM	-0.578***	(0.195)	0.192	(0.352)

Note: Estimates of higher education outcomes as specified in section A.1. HE Level = level of higher education (0-2). Bootstrap standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

A.4 Other tables

Table A7: High school program and level and major college degree: summary statistics

	Academic level higher education			**Major**
	University	Academic college	Professional college	STEM
Study program				
All	*12.4*	*6.1*	*25.5*	*16.9*
Academic				
clas+math	66.5	14.8	12.9	51.0
clas	48.9	10.4	31.2	20.5
math	33.6	18.2	36.4	44.9
other	9.5	6.8	57.8	14.5
Middle-Theoretical				
math	7.2	20.8	44.8	55.2
other	0.1	3.3	43.9	17.1
Middle-Practical	*0.2*	*2.9*	*24.3*	*12.2*
Vocational (13th grade)	*0*	*0.2*	*2.5*	*0.3*

Note: Percentage of all students (also drop outs), conditional on high school program. Three types of higher education options in decreasing order of academic level: university, academic college, professional college. Graduation rates adds up to the total graduation rate of 44.0%. Each level has different programs, aggregated to STEM and other majors, only STEM major is reported. Clas= classical languages included. Math= intensive math.

Table A8: Exclusions because of certificates (in % of certificates)

	Tracks excluded				**Only elective courses excluded**
Current track	Academic	+Middle-Theoretical	+Middle-Practical	+Vocational	
Academic					
grade 7+8	8.9	4.1	4.0	0.8	2.4
grade 9+10	7.9	4.5	4.4	3.8	2.5
grade 11+12	6.2	6.2	6.2	6.2	0
Middle-Theoretical					
grade 7+8	29.1	21.5	19.6	1.2	0.5
grade 9+10	100	16.5	12.5	6.4	0.1
grade 11+12	100	11.2	11.2	11.2	0
Middle-Practical					
grade 7+8	41.5	33.7	30.4	3.8	0.4
grade 9+10	100	100	22.5	9.3	0
grade 11+12	100	100	15.0	15.0	0
Vocational					
grade 7+8	100	100	100	7.1	0
grade 9+10	100	100	100	13.7	0
grade 11+12+13	100	100	100	13.6	0

Note: Summary of implications of A-, B- and C-certificates. C-certificate: repeat grade, i.e. all tracks excluded, B-certificate can exclude entire tracks or only elective courses. Only electives excl. = math options or classical languages excluded by certificate. Upward mobility always excluded from grade 7 on in the vocational track and from grade 9 on in the other tracks. Track switching from grade 11 on is not possible.

Table A9: Transition matrix (in % of students)

	Last choice						
	Acad-clas	Acad-no clas	Middle-theo	Middle-prac	Vocational	Dropout	
First choice							
Acad-clas	11.2	15.8	4.5	2.4	1.5	1.5	36.9
Acad-no clas		9.9	6.0	4.7	3.0	2.4	26.0
Middle-theo		1.3	4.4	3.8	6.4	3.6	19.4
Middle-prac		0.1	0.9	1.0	3.0	1.9	6.9
Vocational					5.5	5.3	10.7
	11.2	27.1	15.8	11.8	19.4	14.6	100.0

Note: Study program choices of students when they enter and leave secondary education.

Table A10: Predictions of the model: counterfactual tracking policies

		Status quo	Policy change B-certificate	
			Repeat	Downgrade
		(%)	(Change in %points)	(Change in %points)
High school				
Academic		*40.4*	0.0 (0.3)	-1.3*** (0.3)
	clas+math	3.3	0.1 (0.0)	0.3*** (0.1)
	clas	3.2	-0.0 (0.1)	0.4*** (0.1)
	math	15.7	0.1 (0.1)	-1.4*** (0.2)
	other	18.3	-0.1 (0.3)	-0.7** (0.3)
Middle-Theoretical		17.0	-1.0*** (0.3)	-1.7*** (0.3)
	math	*2.8*	-0.4*** (0.1)	-0.0 (0.1)
	other	14.2	-0.6** (0.3)	-1.7*** (0.3)
Middle-Practical		7.8	-1.1*** (0.3)	-0.5* (0.3)
Vocational		20.8	-1.2*** (0.3)	4.9*** (0.3)
Dropout		*13.9*	3.3*** (0.3)	-1.4*** (0.3)
Students with at least 1 B-certificate		36.8	-10.7*** (0.6)	-3.8*** (0.5)
Students with at least 1 C-certificate		*30.9*	-0.2 (0.2)	-2.2*** (0.3)
Students with at least 1 year of study delay		34.4	8.1*** (0.5)	-10.7*** (0.7)
Higher education				
Enrollment		*59.4*	-1.7*** (0.3)	-1.7*** (0.3)
Graduation		44.9	-1.8*** (0.3)	-0.3 (0.3)
University degree		11.4	-0.1 (0.1)	0.1 (0.1)
Academic college degree		6.2	-0.2*** (0.0)	-0.0 (0.1)
Professional college degree		27.3	-1.5*** (0.2)	-0.4 (0.2)
Degree in STEM major		16.9	-0.7*** (0.1)	0.0 (0.2)

Note: Predictions from a dynamic model in which students choose study program and distribution of performance. B-certificate = students acquired skills to proceed to next grade but only in track of lower academic level or if they drop elective course. Status quo = students can choose to downgrade or repeat grade after obtaining B-certificate, Repeat = students must repeat grade after obtaining B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining B-certificate. Bootstrap standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).