WORKING PAPERS

N° TSE –1002

March 2019

"A dynamic model of effort choice in high school"

Olivier De Groote

Toulouse School of Economics

A dynamic model of effort choice in high school

Olivier De Groote*

March 28, 2019

Abstract

I estimate a dynamic model of educational decisions that allows for observed and unobserved differences in initial ability. Each year students choose their level of effort by deciding over the academic level of their study program and the likelihood of end-of-year performance. Good performance is costly, but necessary to continue in the program. This replaces traditional approaches, which assume performance follows an exogenous law of motion. I use the model to investigate high school tracking policies and obtain the following results: (1) encouraging underperforming students to switch to less academic programs substantially reduces grade retention and dropout, (2) the resulting decrease in the number of college graduates is small and insignificant, and (3) a model that assumes performance is exogenous ignores a change in unobserved study effort, leading to large biases and falsely concluding there would be an important negative impact on graduation rates in higher education.

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, Thierry Magnac, 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, CEPR Applied IO 2018 and the ES 2018 European Winter Meeting. 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. 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. 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 dropout. These outcomes do not only generate large costs for students, but they also cause negative externalities on society. Therefore, I investigate how to design policies that help in matching students to a study program. This is a general concern in the design of educational systems, but it is especially important in countries that differentiate students already at the age of 10 to 12.[2]

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 careers. I use rich

[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]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).

micro-data of Flanders, the largest region of Belgium. Study programs consist of tracks and elective courses within each track. Students choose a program when they enter high school at age 12. This choice can be updated almost every year after. There is a tracking policy that offers underperforming students the choice between switching out of an academically rigorous program or repeating the grade. I study the impact of study programs that differ in their academic level on higher education enrollment and graduation 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 yearly decisions about their level of academic effort. They do this by choosing among study programs of different level and by deciding on the likelihood of performance at the end of the year. 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 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 obtain a good performance outcome. Good performance is costly, but required to continue in the program and, at the end of high school, to graduate.

In the traditional dynamic discrete choice model, performance would enter as a law of motion that characterizes the state variables in the next period. This law of motion can be estimated and 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, i.e. they choose the probability to be successful. This innovation does not require additional data (like a measure of study effort) or exclusion restrictions. Instead, I make use of a first-order condition that equates the (unobserved) marginal cost of improving the distribution of performance to the marginal benefit it is expected to generate in the future. The benefit of a better performance outcome (i.e. a better state) is easy to derive from a dynamic model. At its optimal level, this has to be consistent with observed program choices and the observed distribution of performance. It therefore allows me to write the marginal costs as a function of the data and other parameters of the model. Standard results for dynamic discrete choice models (Magnac and Thesmar, 2002) can then be used for identification. I also show that the model can be estimated without solving it. Optimal levels of effort within each possible program can be obtained in a first stage using only performance data. I can then apply CCP (Conditional Choice Probability) estimation (Hotz and Miller, 1993) with finite dependence (Arcidiacono and Miller, 2011) to also avoid solving for optimal program choices.

To look at long run effects on educational outcomes, I simultaneously estimate the causal effect of high school study program and grade retention on enrollment in and graduation from higher education. Because of unobserved ability, it is crucial to allow for correlation between the unobservables that play a role in explaining choices and outcomes in both secondary and higher 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 most students make choices and obtain performance outcomes during at least six years. Data on initial cognitive ability and socioeconomic status also allows me to include rich patterns of heterogeneity, even with a small number of unobserved types.

I find a positive impact on obtaining a higher education degree after attending academic programs, and a negative impact of past grade retention. Counterfactual simulations are therefore needed to predict the impact of policies that are expected to change both. Allowing underperforming students to switch to a program of lower academic level (downgrade) as an alternative for repeating a grade has important benefits in the long run. Without this possibility, the number of students with grade retention would increase by about 29%, high school dropout by 28% and the number of college graduates would decrease by 4%. A second counterfactual shows that this policy can be improved. Prohibiting students to repeat a grade (if they have the option to downgrade), would decrease the number of retained students by 30% and dropout by 11%. Enrollment in higher education would go down but the number of students that obtain a higher education degree would not decrease significantly. A welfare analysis shows that this policy can create a loss for students, but this loss is largely offset by the taxpayers' gains. The large impact of initial conditions on effort costs suggests these gains should be invested in early childhood education.

I compare this to the predictions of a traditional dynamic model where students can choose their study program, but performance is exogenous and find an underestimation of the positive effects of both simulations. This is because students react to the policy by improving their performance (i.e. increasing study effort). Therefore, they become less likely to repeat grades or dropout, and more likely to graduate from an academic program. This in turn has an impact on higher education outcomes. For example, in the policy where underperforming students are no longer allowed to repeat a grade, a traditional model would predict an important reduction in the number of college graduates (2.5%) while a model that allows for changes in study effort finds a much smaller and insignificant effect of less than 1%.

This paper contributes to three strands of literature. First, I contribute to the estimation of dynamic discrete choice models in general, and educational decision making in particular. Since the seminal contribution of Keane and Wolpin (1997), dynamic discrete choice models have often been used to evaluate the impact of counterfactual policies on educational decisions. This includes the decision to stay in high school, go to college or choose a major.[3] Allowing students to be forward looking is important because they are expected to react in advance to the future impact of a new policy. In the traditional model, the only channel through which students can respond 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)). Therefore, recent papers have included observable measures of study 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. Therefore, I propose an alternative strategy that can be used when only program and performance data are available. This methodological contribution 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.[4]

A second strand of literature investigates the causal impact of high school on long run outcomes. Altonji *et al.* (2012) reviews the literature on the effects of high school curriculum, initiated by Altonji (1995). 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)). A separate literature looks at the impact of grade retention (Jacob and Lefgren (2009); Manacorda (2012), Cockx *et al.* (2018b)). I

[3] 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 educational decisions (Arcidiacono (2004), Beffy *et al.* (2012)).

[4] There is also a related literature on models of job search in which the probability to find a job (search intensity) is a choice variable (Paserman (2008), van den Berg *et al.* (2015) and Cockx *et al.* (2018a)).

contribute to this literature by jointly analysing high school program choice and grade retention within a structural model. I distinguish between high school programs of different academic level (tracks) and look at differences within tracks, based on math-intensity and the inclusion of classical languages in the curriculum.

Finally, I contribute to the literature on 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 specific groups of students (Guyon *et al.* (2012), Dustmann *et al.* (2017)). Cockx *et al.* (2018b) look at average treatment effects for outcomes within high school for students that are forced to repeat grades or switch tracks. Recent evidence also shows that switching track can diminish negative consequences of early track choice (Dustmann *et al.* (2017), De Groote and Declercq (2018)). I contribute to this literature by investigating how tracking policies during secondary education can help underperforming students to switch to the right track.

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 identification and estimation. I discuss the estimation results in section 5 and I simulate tracking policies in section 6. Finally, I conclude in section 7.

2 Institutional background and data

This section describes the institutional context in Flanders (Belgium) and introduces the data. I make use of the LOSO dataset in which I follow a sample of 5,158 students that started secondary education in 1990.[5] Students were actively followed during high school and therefore the data contains many individual characteristics,

[5]The LOSO data were collected by 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 throughout the paper I discuss the data for this sample of 5,158 students, which covers 80% of the original sample. In the data appendix I discuss in detail why some observations were dropped. This is mainly because some variables were missing, but also because students made choices that were not consistent with the tracking systems as explained in this paper.

choices, performance outcomes and test scores. Afterwards, the students were asked to respond to surveys about their future educational and labor market outcomes which provides information about their higher education career. Details about the both the data and the context are omitted from this text but discussed in the appendix section A.2.

2.1 Study programs

After finishing six grades in elementary school, students enroll in 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. After obtaining a high school degree, students can enroll in higher education.

In full time education, they choose between different high school programs, grouped into tracks that differ in their academic level. 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. Therefore, I follow Cockx *et al.* (2018b) and distinguish between a track preparing mainly for higher education programs (middle-theoretical), and a track that prepares more for the labor market (middle-practical). Students can also choose for the vocational track. This track prepares them for specific occupations that do not require a higher education degree. Within each track, there is a choice between several programs. I aggregate them up to eight study programs. I split up the academic track into four programs: classical languages, intensive math, intensive math + classical languages and other. The middle-theoretical track is split between intensive math and other. This aggregation still allows for a substantial number of students in each program and corresponds to important differences in enrollment and success rates in higher education (Declercq and Verboven, 2015).

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. Leaving in the vocational track after grade 12 is also not considered dropout as students still obtain a certificate that is valued on the labor market (but they need the 13th grade to have access to higher education). Compulsory education laws require students to pursue education until June 30th of the year they reach the age of 18. From the age of 15, they can also decide to leave full time education and start a part-time program in which work and school can be combined.

Although each track prepares for different options after high school, enrollment in almost any higher education option is free of selection by track or by the colleges 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. Similar to Declercq and Verboven (2015, 2018), I distinguish between three levels of higher education (professional college, academic college and university), allow for STEM and non-STEM majors, and for universities, I distinguish between different campuses in Flanders.

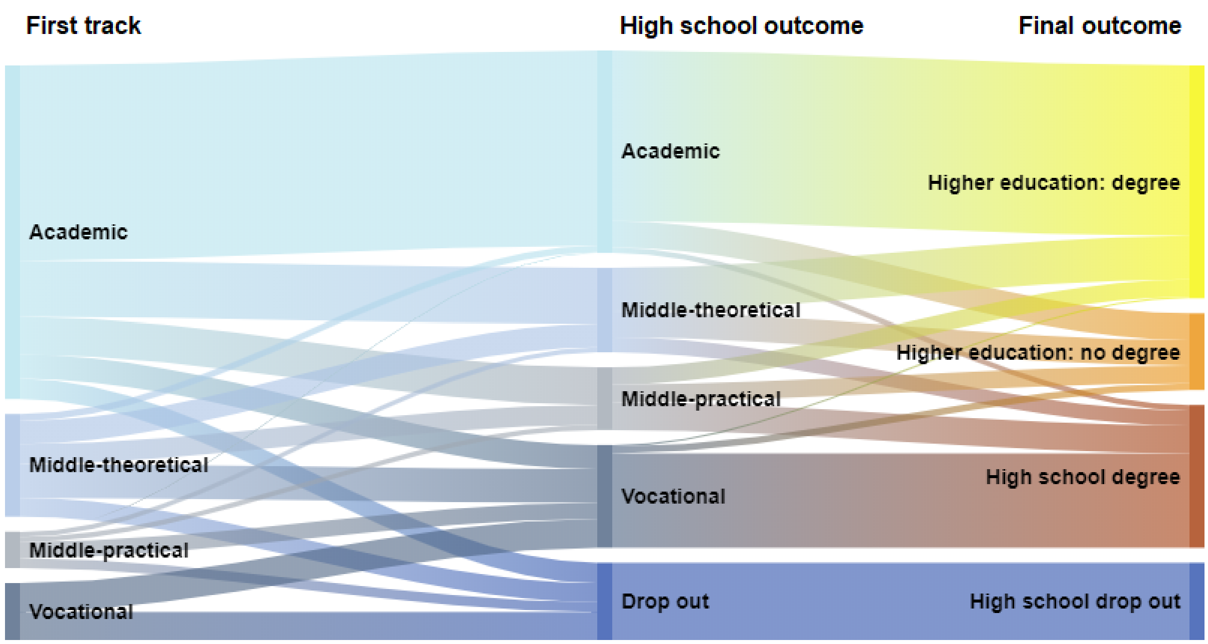
2.2 Mobility

At the start of secondary education, all programs are available. The choice set in the future depends on the past 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 if they did not choose it from the start. Math-intensive programs are available from grade 9 on and similar restrictions apply from then on. Finally, there can be no more switching between full-time programs from grade 11 on.[6]

[6]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. Cockx *et al.* (2018b) apply a similar set of (informal) rules. In the data appendix I discuss the data cleaning and more details about the rules. This shows that only a small number of observations have to be dropped because they are inconsistent with this description.

The restrictions imply that a wrong choice at age 12 can have large consequences for educational attainment. As there can be uncertainty about performance and future preferences, many students like to keep their options open by choosing the academic track in the beginning and gradually move towards their final program. Figure 1 summarizes both the movements within high school (left side) and between high school and the final educational outcome (right side).[7] Most students start in the

Figure 1: Transitions in educational system



Note: Left: program chosen in grade 7, middle: last choice before leaving secondary education, right: final educational outcome. See appendix Table A6 for data on these transitions. Students in the vocational track only obtain a full high school degree that gives access to higher education after an additional 13th grade. They obtain another type of degree after grade 12 and are therefore not considered dropouts if they leave before grade 13. Figure created using Google Charts.

academic track but many transition to another one. There are similar movements from the other tracks but almost always in a downward fashion. Nevertheless, students that move down do not necessarily give up on obtaining a higher education degree. While it is very common for students that graduated from the academic track to

[7] A more detailed overview of transitions to higher education can be found in the Appendix Table A2 and Table A3. As in Declercq and Verboven (2018), I define a degree as three successful years of higher education in a time span of six years.

obtain a higher education degree, most students in the middle-theoretical track and some in the middle-practical track also obtain one. For students that graduated from the vocational track this is very uncommon. Many students that started in the vocational track even drop out of high school, which excludes the possibility to go to higher education.

2.3 Performance and the tracking policy

The transitions in high school are not always a smooth or voluntary process. Each study program comes with its own performance standards. Teachers are expected to keep the quality standards within the program at a certain level. This is done by handing out a certificate to students, based on their performance during the year. An A-certificate is given to students that did not fail a single course. They can move to the next grade and continue in the program. If they failed on some courses, teachers need to decide on the certificate. This can still be an A-certificate, e.g. for students that 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 not in every program. 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 A4). Most of the time students obtain an A-certificate. 7.1% of the certificates are B-certificates and 6.6% are C-certificates. A C-certificate always leads to grade retention if students 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 A5 summarizes the number of students that obtain a B- or C-certificate or accumulate study delay. It then compares their educational outcomes with that of the average student. 32% of students leave high school with at least one year of study delay. These students are 22 %points less likely to enroll in and 24 %points less likely to graduate from higher education than students that were not retained. Part of this is also explained by the higher dropout rates in high school.

2.4 Implications for optimal policies and the required model

Figure 1 shows large differences between higher education outcomes when we compare different tracks. Appendix Table A2 shows that there are also differences between study programs of the same track. Therefore, a policy that encourages students to choose higher levels of academic effort in high school is 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 and is also very costly for society.

These effects can be driven by observed and unobserved initial conditions that could have a direct impact on higher education outcomes. Appendix Table A1 shows the differences between student characteristics and their final program in high school. The dataset contains measures of cognitive ability (language and math), gender and socioeconomic status (SES). The latter is defined as a dummy equal to one if at least one of the parents has completed higher education. We see that academic programs attract mostly female, high ability students with a high SES. In the model I will control for these characteristics and also allow for an unobserved type to capture other characteristics (e.g. noncognitive ability).

Even if the effects are not all driven by initial conditions, 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. Therefore, the model should be able to 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. I do this 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 a new policy that was announced and is close to what will be implemented in Flanders in 2020. It simulates the impact of not allowing students with a B-certificate to repeat the grade.[8] This policy follows from a concern that grade retention is too high and costly for society. However, a potential threat is that this policy might reduce the number of students in tracks that prepare for higher education and thereby decrease the number of college graduates in society.

3 The model

This section introduces a dynamic model of educational choices in which students make yearly decisions about their 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 specify the distribution of end-of year performance. I explicitly model these two yearly decisions using a structural model for the time students are attending high school. I will also model their decision in the period

[8]In the actual implementation of the policy, it will still be possible in some cases to repeat the grade but only if students get their teachers' explicit permission (see answer by the Flemish minister in parliament at 4 October 2018 on question 2410 in period 2017-2018).

they leave high school to capture the expected value of college enrollment. College graduation is modeled as an exogenous process, depending on the outcomes of the structural model. As I explain further in this section, this allows me to be 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 (se) or higher education (he) and $j = 0$ is an outside option, i.e. not attending school. After each year t in high school, students obtain a performance outcome g_{it+1}. Together with the program choice at time t, this 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 how students trade-off the current utility of a study program with its future impact. Next, I discuss the current value by specifying a utility function that includes a fixed cost, independent of performance, and a variable cost that increases in expected performance. I explain how I can connect the model of high school choices to higher education outcomes and I finish this section by explaining how the model is solved. Some details that are specific for the application are left out of the main text and discussed in the appendix section A.4.

3.1 End-of-year performance

Define a variable effort component $y_{it} \in (0, +\infty)$ to be a scalar that influences the distribution of a discrete end-of-year performance outcome g_{it+1}. The lowest outcome (1) does not allow any track in the next grade. 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 explain how this is added to the model in appendix section A.4.

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

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

where $\bar{\eta}^{\bar{g}}_{j,grade}$ denotes the program-grade-specific threshold to obtain at least outcome $\bar{g}$. At time t, students know (and choose) y_{it}, but they do not know the realization of g_{it+1} because of the shock η_{it+1}. This captures uncertainty in grading standards or unexpected events during the year. The information students do have at time t is the probability of obtaining an outcome $\bar{g}$ in program j:

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

with $F(.)$ the cumulative distribution function of the shock. Setting $\bar{\eta}^{1}_{j,grade} = -\infty$ and $\bar{\eta}^{G+1}_{j,grade} = +\infty$ guarantees that all probabilities add up to 1. I assume η_{it+1} is logistically distributed such that

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

Since (1) remains equivalent when adding or subtracting the same term on all sides, I normalize one of the thresholds $\bar{\eta}^{2}_{j,grade} = 0$. 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:

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

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. Therefore, y should 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 by changing the effect of y on performance, but through differences in costs and benefits of the level of y.

3.2 Current and future value of a study program

Each year, students solve a dynamic problem.[9] They choose the study program j with the highest expected lifetime utility, in which the variable effort component y is chosen to maximize the value of the program. The value of each program can be represented by a Bellman equation:

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

with $v_{ijt}(x_{it}, \nu_i, y_{it})$ the conditional value function for student i with observed state variable x and unobserved type ν of choosing program j and variable effort y at time t and $\bar{V}_{t+1}(x_{it+1}, \nu_i) \equiv \int V_{t+1}(x_{it+1}, \nu_i, \varepsilon_{it+1}) h(\varepsilon_{it+1}) d\varepsilon_{it+1}$. As in Rust (1987), this implies that students do not know future realizations of taste shocks, but they know

[9] The decision will often be a joint 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.

the distribution $h(\varepsilon_{it})$, which I assume to be iid extreme value type 1. 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.

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. 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 the future. Since the performance measure g is the only stochastic element in x, the weights are simply the ordered logit probabilities of the performance outcome: $P_j(g_{it+1}|y_{it})$.

3.3 The utility function

As in Keane and Wolpin (1997), I interpret (-) flow utility as an effort cost of going to school. They 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.[10] Since there is no link between costs of schooling and the performance distribution, this excludes any effect of counterfactual simulations through a direct effect on the performance distribution, e.g. because of a change

[10]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).

in unobserved study effort.

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 independent of performance. This captures a distaste to go to school because of differences in preferences or social norms, but also differences in travel time to school. The cost can be negative because students might enjoy going to school or parents can reward (or force) them to go.[11] $c_j(x_{it}, \nu_i)$ is the marginal cost of effort, or the cost of increasing the variable effort component y by 1 unit. The marginal cost captures that, conditional on future values, students dislike the effort that is required to perform better.[12]

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 gets very low (which it is for most students in the data), it becomes costlier to further decrease it.

[11]More generally, it captures any deviation of choices at time t from optimal long run behavior. This means that counterfactual predictions of educational outcomes will be consistent with certain forms or irrational or myopic behavior, as long as these deviations are policy-invariant.

[12]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.

3.4 After secondary education

I assume leaving secondary education is a terminal action, i.e. a student never returns to high school. They either leave the education system, or (if they obtained a high school degree) they choose one of the college options. To avoid making assumptions on how students expect their wages and college success to evolve, I directly parameterize the expected lifetime utility of college enrollment instead of deriving it from a model. In particular I assume the conditional value functions for options outside of high school take the following form:

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

with T_i^{SE} the last period student i spends in high school, $\text{Degree}'_{it} \subset x_{it}$ a vector of dummy variables for the different types of high school degrees a student can obtain, μ^{degree} a vector of parameters to estimate and $\Psi_j^{HEE}(.)$ a function of the state variables that predicts the higher education enrollment (HEE) decision.[13] Since there is no future value term in this conditional value function, these parameters should be interpreted as the total expected lifetime utility from enrolling in option j. If the student obtained a high school degree, j is a specific college option or an outside option of which the utility, net of the value of a degree, is normalized: $\Psi_0^{HEE} = 0$. I distinguish between different locations of colleges, three different levels and two majors (STEM and non-STEM). If the student did not obtain a high school degree, he can only obtain the value of $j = 0$. By normalizing $\Psi_0^{HEE} = 0$, all cost parameters in high school should be interpreted as the one period change with respect to the lifetime value of leaving high school without a degree. Note that this includes potential wages for high school dropouts. Therefore, a high cost of schooling can also be interpreted as an opportunity cost.

[13] 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 conditional logit model with $\Psi_j^{HED}(x_{it}, \nu_i)$ the estimated index that predicts graduation in each campus-level-major combination, conditional on student characteristics, high school program, study delay and the enrollment decision. Note that I did not make any assumptions on students' expectation of higher education success. I simply estimate parameters for $\Psi_j^{HEE}(.)$ to capture the expected lifetime utility of enrolling in each college option. The reason I can do this is because the counterfactual simulations of this paper will only change the high school system, not the higher education system. Therefore, it will result in different high school outcomes (program, dropout, years of study delay), leading to differences in higher education outcomes, but not to differences in the mapping between high school and higher education outcomes. The parameters of $\Psi_j^{HEE}(.)$ and $\Psi_j^{HED}(.)$ are policy-invariant and can be used in counterfactuals. This part of the model is similar to dynamic treatment effect models where behavioral assumptions can be avoided, while still looking at the causal impact of a counterfactual (Heckman et al., 2016).

3.5 Solving the model

I assume it is no longer possible to go to secondary education in $T^{\max} = 10$ such that the model can be solved backwards. Because of the extreme value assumption on the taste shocks ε_{ijt}, I can write the expected value of lifetime utility in the period where secondary education is no longer allowed, using the logsum formula:

$$\bar{V}_{t+1}(x_{it+1}, \nu_i) = \gamma + \ln \sum_{j \in \Phi(x_{it+1})} \exp(\text{Degree}'_{it+1}\mu^{\text{degree}} + \Psi_j^{HEE}(x_{it+1}, \nu_i)) \text{ if } t+1 = T^{\max} \tag{7}$$

with $\gamma \approx 0.577$ the Euler constant and $\Phi_{it+1} = \Phi(x_{it+1})$ the choice set. $\bar{V}_{t+1}$ is used as an input in t (see (4)). First, students look for the optimal value of their 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}. Because $y \in (0, +\infty)$, an interior solution is required and the following first-order condition should be satisfied[14]:

$$c_j(x_{it}, \nu_i) = \beta \sum_{\bar{g}} \frac{\partial P_j(g_{it+1} = \bar{g}|y_{it})}{\partial y_{it}} \bar{V}_{t+1}(x_{it+1}(\bar{g}), \nu_i) \text{ if } y_{it} = y^*_{ijt}. \tag{8}$$

This condition equalizes marginal costs and (expected) marginal benefits. As this does not depend on taste shocks ε or performance shocks η, it implies that students with the same state vector (x_{it}, ν_i) at time t will choose the same effort levels in a given program: $y^*_{ijt} = y^*_{jt}(x_{it}, \nu_i)$. For a binary performance measure, optimal effort in each program has a simple analytic solution:

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

y in period t increases in the discounted benefits of obtaining performance level 2 instead of 1 in period $t+1$, and decreases in marginal costs c in period t. This shows a clear dynamic trade-off. Extra effort at time t is costly but generates benefits in $t+1$. We see that requiring an interior solution 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 A.3).[15] A sufficient condition that is standard in dynamic discrete choice models is to assume that a student always believes there is a non-zero probability of avoiding the worst performance outcome.

When students know the optimal levels of effort in each program, they can choose

[14] I relax this in counterfactual simulations by checking if a corner solution ($y = 0$) is optimal. This can be done using a grid search for y. This is necessary because certain policy changes no longer make it obvious that a higher performance outcome is always preferred and a policy could also change the marginal benefits in such a way that it makes some students no longer willing to pay any cost for y.

[15] In the appendix I also 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.

the program with the highest value of $v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) + \varepsilon_{ijt}$. This results in the following logit choice probabilities:

$$\Pr(d^j_{it} = 1 | x_{it}, \nu_i) = \frac{\exp(v_{ijt}(x_{it}, \nu_i, y^*_{ijt}))}{\sum_{j' \in \Phi(x_{it})} \exp(v_{ij't}(x_{it}, \nu_i, y^*_{ij't}))} \tag{9}$$

with d^j_{it} an indicator for choosing j, v_{ijt} given by (4) for $j \in se$ and (6) for $j \in \{0, he\}$. $\bar{V}_t$ can also be calculated using:

$$\bar{V}_t(x_{it}, \nu_i) = \gamma + \ln \sum_{j \in \Phi(x_{it})} \exp(v_{ijt}(x_{it}, \nu_i, y^*_{ijt})). \tag{10}$$

These steps can be repeated until the first period to solve the entire model.

4 Identification and estimation

The primitives we need to identify that depend on (x_{it}, ν_i) are the fixed costs $C^0_j(.)$, marginal costs $c_j(.)$, the expected value of college enrollment $\Psi^{HEE}_j(.)$ and the parameters that predict higher education graduation $\Psi^{HED}_j(.)$. We also need to identify the thresholds for performance outcomes in each track and grade: $\bar{\eta}_{j,grade} = \{\bar{\eta}^1_{j,grade}, ..., \bar{\eta}^G_{j,grade}\}$, the value of obtaining a high school degree μ^{degree} and the unobserved type distribution.

I first discuss identification, assuming the econometrician observes the type ν of each student. I explain how higher education data identifies some outcomes of the model and how high school data can be used to separately identify fixed costs, marginal costs and the value of a degree. Subsequently I show how Hotz and Miller's (1993) CCP method can be applied to avoid solving the model during estimation. Then, I explain how to allow for unobserved types. I will be flexible on the functional form assumptions of all aforementioned components to show that they are nonparametrically identified. However, for the applications I will use a parametric estimation approach that I summarize at the end of this section.

4.1 Higher education data

As we are only interested in counterfactual simulations during high school, the model only attempts to explain their impact (contained in x) on higher education. This means that we do distinguish between an impact on effort costs, information, wage returns or any other channel. Therefore, the parameters related to higher education capture a reduced form effect and are straightforward to identify from higher education data alone. With the extreme value type 1 assumption on the error terms, enrollment parameters are identified by log odds ratios in the data for students who just left high school (with a degree):

$$\ln\left(\frac{\Pr(d_{it}^{j}=1|x_{it},\nu_i)}{\Pr(d_{it}^{0}=1|x_{it},\nu_i)}\right)=\Psi_j^{HEE}(x_{it},\nu_i) \text{ for } t=T_i^{SE}+1.$$

Similarly, identification of $\Psi_j^{HED}(.)$ uses probabilities of the programs in which students graduate and also takes into account the enrollment decision in x.

4.2 High school data

The other primitives are identified using data on performance and study programs during high school. As for most dynamic discrete choice models, identification of the flow utility depends on the distributional assumptions on taste shocks ε_{ijt}, the flow utility of one option has to be normalized and the discount factor β is set before estimation (Magnac and Thesmar, 2002; Abbring and Daljord, 2018).[16] The novelty 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 and a variable cost that depends on an estimated marginal cost. Therefore, instead of estimating a flow utility and a law of motion, I estimate two components of the utility function. I do this by giving a structural interpretation to the performance outcomes. To have a scalar choice variable (y) that affects a discrete outcome, I also assume the ordered logit process

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

on the performance outcome specified in section 3.1.[17]

I first give an example of how the two types of data identify different primitives and then show this more formally.

4.2.1 Example

Take students that are in high school in the last grade of the same study program and in the final year where this is possible. If they perform well, they obtain a high school degree and can choose between enrolling in different college options or dropping out of school. If they fail, they can only drop out. Say now their characteristics are different but, conditional on obtaining a high school degree, they predict the same choice in the next period. This implies that they have the same marginal benefits (as a function of y) at time t. If we now observe a different probability to obtain a high school degree, i.e. a different distribution of performance g_{it+1} in the same program, it can only be explained by differences in marginal costs because the first-order condition would otherwise not be satisfied. A traditional model would estimate parameters of a law of motion, without taking into account the incentives of students. Assume instead they do have the same distribution of performance, but they have different program choice probabilities at time t. In this case the fixed cost $C_j^0(x_{it}, \nu_i)$ are the only way to rationalize this. A traditional model would estimate a different flow utility (i.e. the sum of fixed and variable costs here).

4.2.2 Performance

The identification strategy for the marginal costs follows from the first-order condition (8). As explained in the previous section, the condition implies that students with the same state vector will choose the same effort levels within each program. We can

[17] This assumption is often used in applications, see e.g. Joensen and Mattana (2017).

substitute $y^*_{jt}(x_{it}, \nu_i)$ in the definition of y (3):

$$y^*_{jt}(x_{it}, \nu_i) = \frac{1 - P_j(g_{it+1} = 1 | y^*_{jt}(x_{it}, \nu_i))}{P_j(g_{it+1} = 1 | y^*_{jt}(x_{it}, \nu_i))}. \tag{11}$$

Note that both x_{it} and ν_i are observed here, therefore $y^*_{jt}(.)$ is easily obtained from the observed probability to obtain the lowest performance outcome because $P_j(g_{it+1} = 1 | y^*_{jt}(x_{it}, \nu_i)) = \Pr(g_{it+1} = 1 | x_{it}, \nu_i, d^j_{it} = 1)$ when students behave optimally in the data. After obtaining $y^*_{jt}(.)$, the probabilities to reach other performance levels can be used to recover the thresholds $\bar{\eta}_{j,grade}$.[18]

A second reason the first-order condition is helpful, is that it allows us to write the conditional value functions without an unknown marginal cost function. Substituting the utility function (5) in the conditional value function (4), after substituting marginal costs by (8) gives:

$$\begin{aligned} & v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) \\ & = -C^0_j(x_{it}, \nu_i) \\ & + \beta \sum_{\bar{g}} \left[\bar{V}_{t+1}(x_{it+1}(\bar{g}), \nu_i) \left(P_j(g_{it+1} = \bar{g} | y^*_{jt}(x_{it}, \nu_i)) - \left. \frac{\partial P_j(g_{it+1} = \bar{g} | y_{it})}{\partial y_{it}} \right|_{y_{it} = y^*_{jt}(x_{it}, \nu_i)} y^*_{jt}(x_{it}, \nu_i) \right) \right]. \end{aligned} \tag{12}$$

Identification of $y^*_{jt}(x_{it}, \nu_i)$ and $P_j(g_{it+1} = 1 | y^*_{jt}(x_{it}, \nu_i))$ is explained above. $\frac{\partial P_j(g_{it} = \bar{g} | y_{it})}{\partial y_{it}}$ is a function of these two components and can be derived from the distributional assumptions on the performance measure (see appendix section A.3).

4.2.3 Study program

After substituting out the marginal cost function, and knowing the parameters that predict enrollment into higher education, conditional value functions during high school still depend on fixed costs parameters and the value of a high school degree (see (12) with $\bar{V}_{t+1}$ given by (7) in the final period and (10) before). Because of the extreme value assumption on the error terms, the program choice probabilities

[18] This can be done by maximum likelihood or by relaxing the structure imposed here and making the thresholds depend on (x_{it}, ν_i).

at the optimal values of y are given by logit probabilities (9). This implies that log odds ratios in the program choice probabilities only identify differences in conditional value functions, which creates a similar problem as in the static literature where only differences in utility are identified. Therefore, I normalize the utility of the outside option $j = 0$ (leaving high school) such that the cost parameters are interpreted as the one period difference with the utility of leaving high school.

Since leaving high school is assumed to be a terminal action, I could set $v_{i0t} = u_{i0t} = 0$. However, in contrast to a static model, normalizing the utility of one option in every state is not innocuous (Kalouptsidi et al., 2016). Here it would imply an assumption that students only exert effort in school to have the possibility to go to higher education, and not for the intrinsic value of a high school degree (e.g. because of social norms but also the direct impact on wages). Therefore, I also estimate the value of a high school degree and set $v_{i0t} = \text{Degree}'_{it}\mu^{\text{degree}}$. As in Eckstein and Wolpin (1999), the value of a degree can be identified from choices in secondary education. In particular, differences in dropout rates between students with low and high chances of obtaining a high school degree help to identify this effect, but exclusion restrictions are needed to separate this from differences in fixed costs. Therefore, I allow distance to college to (endogenously) affect performance in high school, but it is excluded from fixed costs.[19]

4.3 Avoid solving for program choices

All primitives are identified but estimation requires solving the model for optimal program choices which makes it computationally intensive. Hotz and Miller (1993) introduced the CCP method as an alternative, which is particularly useful if there is a terminal action (Arcidiacono and Ellickson, 2011). In this section I show that it

[19] Also the parametric assumptions I discuss in the appendix section A.4 help for identification. In particular, I restrict the change in fixed costs over grades to be linear in each track. This helps identification here because higher grades also imply a decrease in uncertainty about eventual graduation and a higher discounted value of the degree. Note that it would also be possible to make μ^{degree} depend on student characteristics if there is sufficient variation in the data.

can be applied in the current model. Hotz and Miller (1993) show that the future value term can be written as the conditional value function of an arbitrary choice and a nonnegative correction term that depends on its probability in the data:

$$\bar{V}_{t+1}(x_{it+1}, \nu_i) = \gamma + v_{id^*t+1}(x_{it+1}, \nu_i) - \ln \Pr(d^*_{it+1}|x_{it+1}, \nu_i) \tag{13}$$

with $\gamma \approx 0.577$ the Euler constant, d^*_{it+1} a 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. 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 (13), with $\Psi_0^{HEE}(.) = 0$:

$$\overline{V}_{t+1}(x_{it+1}, \nu_i) = \gamma + \text{Degree}'_{it}\mu^{\text{degree}} - \ln \Pr(d^0_{it+1} = 1|x_{it+1}, \nu_i). \tag{14}$$

We can now substitute (14) in (12), such that for all $j \in se$:

$$\begin{aligned}
&v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) \\
&= -C^0_j(x_{it}, \nu_i) + \beta\gamma \\
&+ \beta \sum_{\bar{g}} \left[\begin{array}{c} \left(\text{Degree}'_{it}(\bar{g})\mu^{\text{degree}} - \ln \Pr(d^0_{it+1} = 1|x_{it+1}(\bar{g}), \nu_i)\right) \\ \left(P_j(g_{it+1} = \bar{g}|y^*_{jt}(x_{it}, \nu_i)) - \frac{\partial P_j(g_{it+1}=\bar{g}|y_{it})}{\partial y_{it}}\Big|_{y_{it}=y^*_{jt}(x_{it},\nu_i)} y^*_{jt}(x_{it}, \nu_i)\right) \end{array} \right].
\end{aligned} \tag{15}$$

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 $\Pr(d^0_{it+1} = 1|x_{it+1}, \nu_i)$ can be recovered from the data before estimating the model.

Because of compulsory education laws, the outside option is not always in the choice set. In the appendix section A.4.5, I show how finite dependence (Arcidiacono and Ellickson, 2011; Arcidiacono and Miller, 2011) can be used to overcome this problem.

4.4 Unobserved types

I allow the type of a student (ν_i) to be unobserved by the econometrician. This can be identified using the panel structure of the data. 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 help in identifying the unobserved type. 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. This instrumental variable strategy helps in separately identifying the unobserved types from the effect of a high school program.[20]

The above steps are no longer straightforward if types are unobserved. Therefore, Arcidiacono and Miller's (2011) adaptation of the EM algorithm is convenient. In sum, it starts by specifying the number of types and letting each student have a random probability to belong to each type. The above strategy to identify all primitives can then be applied by using the type probabilities as weights. The joint likelihood of program choices, performance and higher education outcomes, conditional on each type, is then used to update the type probabilities using Bayes rule. This is repeated until convergence.

4.5 Estimation approach in finite samples

I showed that identification is possible without strong functional form assumptions. Some exclusion restrictions where introduced to help identification of the value of a degree and the unobserved types, but this would be similar in a traditional dynamic

[20]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.

model where students are not able to choose the distribution of performance. The shows that identification is coming from replacing the assumption that performance is exogenous to assuming students optimally choose its distribution. In practice, estimation cannot proceed without further structure because of the limited number of observations in each realization of the state. Therefore, I propose the following estimation algorithm. The joint likelihood function of this estimator and further details can be found in the appendix section A.4.7.

I assume there are two unobserved types in the population. I start by assigning a random probability to each student and use them as weights in what follows.

4.5.1 Higher education data

I propose functional forms for $\Psi_j^{HEE}(.)$ and $\Psi_j^{HED}(.)$ and estimate them as parameters of a conditional logit, using maximum likelihood.

4.5.2 High school data

I recover the optimal levels of the variable effort component and the performance thresholds by estimating an ordered logit model with index $\ln y^*_{jt}(x_{it}, \nu_i)$ and cut points $\bar{\eta}_{j,grade}$. The index is specified as a parametric, but flexible, function of (x_{it}, ν_i). As in Arcidiacono *et al.* (2016), I also obtain predicted values of $\Pr(d_{it} = 1|x_{it}, \nu_i)$ (the CCPs) by estimating a flexible conditional logit with a similar index.

4.5.3 Update types

I use the predicted values of CCPs, performance and higher education outcomes by type to update the individual type probabilities using Bayes rule. I then repeat the previous steps until convergence of the joint likelihood of the data. This is an application of Arcidiacono and Miller's (2011) two-step approach which does not require the structural cost parameters to find the types because it is using CCPs to calculate the joint likelihood function.

4.5.4 Cost estimates

After convergence of the joint likelihood function, I use the logit probabilities (9) with the CCP-representation of the conditional value functions (12) to estimate the value of a degree μ^{degree} and a specification for fixed costs $C_j^0(.)$ using maximum likelihood with the type probabilities as weights. Finally, marginal costs $c_j(.)$ can be recovered from the first-order condition (8) without imposing any additional structure. Standard errors are obtained using a bootstrap procedure.[21]

5 Estimation results

This section discusses the structural schooling cost estimates and the estimates of higher education outcomes. To check the fit of the model and to run counterfactual simulations, I use these estimates to 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. The model fit and details about the simulations are explained in the Appendix section A.5. The model does a decent job in capturing the patterns in the data such that it can be used for the counterfactual simulations in the next section.

5.1 Effort costs

I focus the discussion on the impact of student characteristics on effort costs using Table 1 and the impact of study delay and switches in Table 2. Appendix Tables A7 and A8 contain the estimates unrelated to student characteristics and the interactions of student characteristics with elective courses. Table A12 shows the intrinsic value of a high school degree.

[21] I sample students with replacement from the observed distribution of the data and use 150 replications. Since the EM algorithm takes some time to converge, I do not correct for estimation

Table 1: Costs of schooling: student characteristics and academic level

	Fixed costs				Log of marginal costs			
	Baseline effect		Interaction with academic level		Baseline effect		Interaction with academic level	
Male	-19.161	(9.714)	17.748	(4.355)	0.740	(0.114)	0.098	(0.061)
Language ability	7.468	(5.485)	-36.328	(5.132)	-0.637	(0.075)	-0.095	(0.088)
Math ability	1.117	(5.081)	-23.197	(4.523)	-0.217	(0.063)	-0.370	(0.051)
High SES	-19.232	(15.555)	-18.782	(5.818)	-0.676	(0.219)	0.016	(0.101)
Type 2	-41.781	(16.422)	85.647	(11.607)	3.265	(0.370)	-0.368	(0.156)

Note: Estimates of a sample of 5,158 students or 33,239 student-year observations. Scale = minutes of daily travel time. The marginal costs in the model are a flexible function of state variables, this table summarizes them by regressing their 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. Level = academic level of high school program (0-3). Bootstrap standard errors in parentheses.

The functional form assumption on the fixed costs is the same as shown in the tables. I divide them by a common fixed cost of travel time such that they can be interpreted in daily minutes of travel time. The marginal costs are a nonlinear function of probabilities in the data and other parameters of the model (see section 4.5). For interpretation purposes only, I perform an OLS regression on the logarithmic transformation of the estimated marginal costs with the same structure as the fixed costs.

We see that male students have a 19-minute lower fixed cost to attend the benchmark vocational program, but the sign of this effect changes for the most academic programs. The marginal cost estimates reveal that they have a harder time to obtain good performance outcomes. For the same increase in expected performance, a male students pays twice (exp(0.740)) the cost of a female students.

Higher cognitive ability leads to decreases in fixed costs (except for the benchmark vocational track). Also marginal costs are strongly affected. An increase by 1% of a standard deviation in language ability leads to a decrease of 0.6% in marginal costs in the vocational track (level=0) and 0.9 in the academic track (level=3).[22] The same

error in the probabilities to belong to each type.

[22] I also estimated a model where academic level is proxied by the hours of academic courses (which

increase in math ability leads to a decrease of 0.2% in the vocational track, but a much larger 1.3% decrease in the academic track.

Despite the controls for cognitive ability, parental background still matters. High SES students are more favorable towards programs of higher academic level and marginal costs are lower. The magnitudes are similar to a standard deviation increase in language ability. We also see that there is still a lot of persistent heterogeneity in the data that observable characteristics are not capturing. Appendix Table A9 shows that 30% of students belong to type 1 and 70% to type 2. Type 1 captures a group that experiences little trouble in successfully completing high school in tracks of high academic level, compared to most students (type 2). One reason is that they have much lower fixed costs when they opt for more academic programs, equivalent to 86 minutes of daily travel time for each step. Also marginal costs are much lower, making it easier to stay in academic programs. A type 1 students only pays 4% to 12% of the marginal cost of a type 2 student.[23]

Table 2 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 a result of demotivation. The same increases in expected performance might be perceived more costly for students with study delay because they might loose interest in studying. On the contrary, we find decreases in marginal costs when students are repeating a grade in programs of high academic level. This can result from the fact that students see the same course material for a second time, making it easier to succeed. At the same time, the fixed costs estimates show that students dislike repeating a grade. This shows a clear trade-off: students dislike repeating a grade, but it does help them to perform well in more academic programs. 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.

varies over both tracks and grades) and obtain similar results.

[23]In the benchmark (=vocational) track this is $\exp(-3.265) = 4\%$, in the academic track it is $\exp(-(3.265 + 3 \times (-0.368))) = 12\%$.

Table 2: Costs of schooling: repeating and switching

	Fixed costs				Log of marginal costs			
	Baseline effect		Interaction with academic level		Baseline effect		Interaction with academic level	
Repeat	274.710	(33.156)	79.903	(13.121)	0.238	(0.194)	-0.461	(0.124)
Study delay	-9.932	(6.884)	10.026	(4.884)	0.637	(0.104)	0.109	(0.064)
Downgrade	169.088	(21.878)			0.174	(0.094)		
Upgrade	325.724	(41.557)			0.262	(0.179)		
Stay in clas	-16.062	(14.652)			0.155	(0.431)		
Stay in math	-203.545	(26.127)			0.653	(0.276)		

Note: Estimates of a sample of 5,158 students or 33,239 student-year observations. Scale = minutes of daily travel time. The marginal costs in the model are a flexible function of state variables, this table summarizes them by regressing their logarithmic transformation on the same variables that enter the fixed costs. Clas= classical languages included. Math= intensive math. Downgrade: switch to lower academic level. Upgrade: switch to higher academic level. Bootstrap standard errors in parentheses.

5.2 Higher education outcomes

All estimates for higher education outcomes can be found in appendix Tables A13, A14 and A15. We see 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 the unobserved type of students, showing that it is 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 calculate the total Average Treatment Effects on the Treated (ATT) of each study program and compare this to a comparison of the raw means in the data (see Table 3). The ATTs are calculated as follows:

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

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 higher education outcome (enrollment or graduation) 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

Table 3: Higher education and high school outcomes: difference in means and ATTs

	Enrollment				Degree			
	Mean diff		ATT		Mean diff		ATT	
Study program								
Academic								
clas+math	5.1	(1.1)	1.7	(0.3)	20.1	(2.3)	8.2	(1.9)
clas	5.2	(0.9)	1.2	(0.2)	16.4	(2.6)	5.4	(2.1)
math	3.7	(1.0)	2.6	(0.3)	14.0	(2.2)	9.6	(1.8)
other	*benchmark*				*benchmark*			
Middle-Theoretical								
math	5.1	(1.1)	4.0	(1.5)	-1.3	(4.2)	5.8	(3.9)
other	-15.2	(1.8)	-5.9	(2.1)	-26.2	(2.6)	-9.5	(2.8)
Middle-Practical	-39.3	(2.2)	-26.4	(2.8)	-46.6	(2.6)	-20.1	(3.2)
Vocational	-80.7	(1.7)	-64.6	(3.3)	-71.5	(1.9)	-37.2	(3.4)
One year of study delay	-26.0	(1.7)	-4.9	(1.1)	-33.9	(1.4)	-12.3	(1.3)
Data	58.2				44.0			

Note: Effects on enrollment 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. Average treatment effects on the treated (ATT) make use of the causal estimates of enrollment and graduation equations. ATTs are calculated using indexes, specified in appendix section A.4, for each individual at the realization of other variables. Effects on obtaining higher education degree are total effects, i.e. they also take into account effects through enrollment. Clas= classical languages included. Math= intensive math. Bootstrap standard errors in parentheses.

the graduation track replaced by an arbitrary benchmark program j^0. The ATT then calculates the average effect on HE of graduating high school in 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 of the entire model but it is a "ceteris paribus" causal effect, i.e. it is the effect of one variable if all other variables that were realized at the time of 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 much smaller extent. I find that graduating from the academic track (without classical languages or extra math) leads to an increase in college graduation of 20 %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 (9.5 %points). The estimates show that elective courses mainly matter for the type of higher education but we also see that overall graduation rates are higher when students had classical languages or intensive math in their program.

For study delay, I find a negative impact on higher education enrollment of 5 %points and an even stronger negative impact if 12 %points on obtaining a higher education degree.

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 they are allowed to transition to the next grade but have to switch to a program of lower academic level or drop an elective course. In this case they obtain a B-certificate. This allows underperforming students to avoid grade retention, but they can still opt for the same program if they are willing to repeat the grade. I compare the current policy to two alternatives:

Counterfactual 1: Repeat

Students are forced to repeat a grade when they obtain a B-certificate. This removes the option to avoid grade retention by switching to a different program if they underperformed this year. It makes the system less flexible and allows us to quantify the importance of the current flexibility.

Counterfactual 2: Downgrade

Students are forced to switch to a different program when they obtain a B-certificate, without repeating the grade. It resembles a policy that will be implemented in Flanders to reduce grade retention.

I first discuss the predicted effect of each policy using the proposed model. I then discuss why the downgrade policy is expected to be beneficial for society, despite the decrease in student welfare. I then discuss heterogeneity and the impact of initial conditions and conclude by comparing my results to those of a model that would assume performance is exogenous. Details about the calculation of welfare effects can be found in the appendix section A.5.2.

6.1 Impact of both policy changes

Table 4 compares the outcomes of the two counterfactuals to the status quo scenario. The "Repeat" policy only shows worse outcomes. It does not manage to significantly increase graduation rates from the academic track, instead we see an increase in the dropout rate of 4 %points, which is a 28% increase in the total number. Not surprisingly, the share of students with grade retention increases by a large amount (9 %points). As a result, enrollment and graduation rates in higher education decrease by 2 %points. I also look at the welfare effects from the student's perspective. Since travel time enters the model, I can look at differences in welfare in kilometers. To facilitate interpretation and compare it to potential government savings, I instead assume an opportunity cost of $10/hour and interpret it in dollars. Student welfare decreases on average by $2,140, which is mainly driven by the increase in fixed costs and can be explained by the impact of repeating grades. Dropout is also decreasing the expected payoff after leaving high school because these students do not obtain a degree and can no longer enroll in college. The increase in grade retention and the decrease in college graduates is also expected to have large negative externalities

Table 4: Counterfactual tracking policy

	Status quo	Policy change B-certificate			
		Repeat		Downgrade	
Panel A: educational outcomes	*%*	*Change in %points*			
High school					
Academic track	40.02	0.17	(0.30)	-1.13	(0.32)
Middle-theoretical track	16.10	-0.96	(0.29)	-1.52	(0.26)
Middle-practical track	8.14	-1.13	(0.24)	-0.77	(0.26)
Vocational track	21.57	-2.01	(0.32)	5.03	(0.29)
Dropout	14.17	3.94	(0.33)	-1.61	(0.25)
At least 1 B-certificate	37.53	-10.23	(0.71)	-3.49	(0.34)
At least 1 C-certificate	30.69	-0.23	(0.33)	-2.15	(0.24)
At least 1 year of study delay	33.22	9.48	(0.57)	-9.82	(0.55)
Higher education					
Enrollment	58.15	-1.76	(0.24)	-1.40	(0.21)
Graduation	44.25	-1.70	(0.22)	-0.30	(0.18)
Panel B: student welfare		*Change in $1000*			
Total student welfare		-2.14	(0.26)	-1.02	(0.14)
Fixed costs (-)		0.85	(0.12)	-0.48	(0.10)
Variable costs (-)		0.49	(0.08)	0.21	(0.03)
Expected payoff after high school (+)		-0.65	(0.10)	0.32	(0.07)
Taste shocks (+)		-0.15	(0.10)	-1.61	(0.19)

Note: Predictions from the dynamic model. C-certificate: repeat grade. B-certificate = students acquired skills to proceed to next grade but only if they downgrade, i.e. switch to track of lower academic level or 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.

that are not considered in this exercise. I conclude that the current flexibility in the tracking policy is better than a strict pass or fail policy.

In the "Downgrade" policy, students who obtained a B-certificate are no longer allowed to repeat the grade. This would lead to a decrease in grade retention rates by 10 %points and dropout rates by 1.6 %points. This does come at a cost in the short run. Students switch to programs of lower academic level, which decreases enrollment rates in higher education by 1.4 %points. However, graduation rates only decrease by an insignificant 0.3 %points (or less than 1% of the total number), which can be explained by the strong effect of study delay on graduation. Since the policy restricts the choice set of students, their welfare unambiguously goes down. On average they lose \$1,020, despite a reduction in the fixed costs of \$480 and an increase in their expected payoff after high school of \$320. This is partly because they increase their study effort in response to the policy, leading to a loss of \$210 in variable effort costs, but the main cost of the policy comes from the reduction in their choice set that makes them miss out on \$1,610, coming from unobserved taste shocks.

I conclude that the current tracking policy is a good way 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 leads to large increases in grade retention and dropout, without an impact on graduation from higher education. A large part of the cost of grade retention is borne by the government. Therefore, it is important to look further into the welfare impact of this policy.

6.2 Externalities of the downgrade policy

Despite the negative impact on student welfare, it is easy to argue that this policy is beneficial for society. First, the average loss of \$1,020 is small compared to the costs of grade retention for society. The OECD estimates the direct cost to the education system of a student that repeats a grade to be \$9,713. Decreasing grade retention rates by 9.82 %points then generates a government saving of \$950 per student. On top of that, gains of a year in tax payments would bring in an additional \$1,960. In the future, we would expect additional externalities that I do not take into account here. In particular, we see a decrease in high school dropout, which is expected to increase tax income and decrease payments through unemployment benefits. At the higher education level, we see a decrease in enrollment and no significant decrease in graduation. Since 90% of higher education spending comes from public sources (OECD, 2012), this increase in efficiency is increasing the positive externality. To avoid strong assumptions, I do not quantify these externalities but it is clear that they would further strengthen the arguments in favor of the policy.

A second reason this policy is expected to be beneficial is because the estimate of the expected loss in student welfare is likely an upper bound for their actual loss. Both taste shocks and fixed costs are assumed to be components of the utility function, i.e. they are not the result of a mistake in their choice but students value them. An alternative interpretation is that they are, at least partly, deviations from the student's optimal path towards deriving value from a high school degree or from attending higher education. Deviations could be explained by myopia, peer pressure, parent's preferences or optimization errors. As long as these deviations are policy-invariant, the counterfactual simulations still hold for educational outcomes, but the welfare impact can be very different, and students might even gain from the policy. Similarly, the fact that enrollment rates in higher education go down while graduation remains the same could be an indication of biased expectations of students about their college success. Therefore, the increase in the actual payoff after high school is likely larger than the increase in expected payoff we account for in the model.

6.3 Heterogeneity and the importance of initial conditions

Even if the average impact on student welfare is small, many students could be hurt by the policy change. It is also interesting to compare the effect of the policy to the impact of initial conditions as savings in high school could be used to target initial conditions directly. To investigate this further, I regress four predicted outcomes of Table 4 on student characteristics in each policy scenario. I look at the impact on obtaining study delay, dropping out of high school, obtaining a higher education degree and student welfare. The effects within the status quo show how the outcomes differ between students with different initial conditions. The interactions with the counterfactuals explains how this changes in the alternative policies. The results can be found in the appendix Table A16.[24]

The impact of initial conditions is large for every outcome. A standard deviation decrease in language ability makes a student 7 %points less likely to obtain study delay, 7 %points less likely to drop out, 15 %points more likely to graduate from college and derive $16,270 more from the current high school system. The impact of math ability, SES or being female are similar in magnitude on most outcomes. Math ability does have a smaller effect on study delay and gender is more important to explain dropout and less important for higher education. The effect of belonging to unobserved type 1 instead of 2 are about twice the size of any of the observable characteristics.

The results put into perspective the impact of the policy change in high school on student outcomes. The "Downgrade" policy is able to achieve large effects in terms of reducing study delay, resulting in large savings on educational spending, but its impact on other outcomes is marginal compared to the impact of initial conditions. This highlights the importance of preparing students before they enter high school, rather than having them exert additional effort (studying harder or repeating grades) during high school. This is consistent with the literature on gains from early childhood education through dynamic complementarities (Heckman and Mosso, 2014; Cunha and Heckman, 2009). High schools should aim at efficiently fostering the skills that students acquired before entering, which is not consistent with letting students repeat grades in order to graduate from more academic programs. The savings that result from this could be used to improve these initial conditions directly (see appendix section A.5.2 for a calculation of the expected effects).

The effect of the policies does not vary a lot over students but there are some small differences in their size. Type 2 students are generally more affected by the

[24]Note that for welfare, only differences are identified. Therefore, the constant is a nuisance parameter and the effect of student background should be interpreted as the effect on student welfare, keeping the utility of the outside good (not attending high school) fixed. This way, it captures changes in fixed and variable effort costs, taste shocks and the expected payoff of leaving high school, but not any effect it might have on the dropout utility.

policy changes. The effects on dropout are more pronounced for male students and high SES and observed ability leads to smaller welfare losses from the repeat policy.

Even though the estimates do not show a lot of heterogeneity in the impact of the policy, it is important to note that these are ex-ante welfare losses, i.e. they are the expected losses by students before they enter high school. Many of them never get a B-certificate and should not be strongly affected (although the risk of getting one is also affecting their behavior). In the appendix Table A17 I compare predictions of the four outcomes, conditional on obtaining a B-certificate in the status quo scenario. We see that most of the effects were driven by students that would have obtained a B-certificate in the status quo as they show large effects on all outcomes. Others are affected too but to a much smaller extent. For example, the welfare loss is 2 ("Downgrade" policy) to 6 times ("Repeat" policy) smaller and the effects on study delay and dropout are almost entirely coming from the group that would have obtained a B-certificate in the status quo.

6.4 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 parameters 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 (unobserved) study effort and thereby (observed) 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 innovation 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 as in the status quo. To estimate this model, I also set the marginal costs in the utility function to 0 such that "fixed cost" estimates capture the entire flow utility. Table 5 compares the counterfactual predictions.

Table 5: Bias in policy simulations with exogenous performance

	Policy change B-certificate							
	Repeat				Downgrade			
Students can adjust study effort	Yes	No	Bias		Yes	No	Bias	
Panel A: educational outcomes	*Change in %points*							
High school								
Academic track	0.17	-0.61	-0.78	(0.19)	-1.13	-1.88	-0.75	(0.18)
Middle-theoretical track	-0.96	-1.32	-0.35	(0.19)	-1.52	-1.88	-0.36	(0.17)
Middle-practical track	-1.13	-1.73	-0.60	(0.17)	-0.77	-1.00	-0.23	(0.16)
Vocational track	-2.01	-1.04	0.96	(0.24)	5.03	6.21	1.19	(0.20)
Dropout	3.94	4.70	0.77	(0.20)	-1.61	-1.46	0.15	(0.12)
At least 1 B-certificate	-10.23	-6.32	3.91	(0.44)	-3.49	-0.86	2.63	(0.29)
At least 1 C-certificate	-0.23	0.22	0.44	(0.20)	-2.15	-1.69	0.45	(0.15)
At least 1 year of study delay	9.48	11.61	2.13	(0.36)	-9.82	-9.19	0.63	(0.27)
Higher education								
Enrollment	-1.76	-3.02	-1.27	(0.17)	-1.40	-2.27	-0.87	(0.11)
Graduation	-1.70	-2.69	-0.99	(0.15)	-0.30	-1.12	-0.81	(0.09)
Panel B: student welfare	*Change in $1000*							
Total student welfare	-2.14	-2.23	-0.09	(0.09)	-1.02	-0.99	0.03	(0.05)
Fixed costs (-)	0.85	1.26	0.41	(0.12)	-0.48	-0.63	-0.14	(0.05)
Variable costs (-)	0.49	0.00	-0.49	(0.08)	0.21	0.00	-0.21	(0.03)
Expected payoff after high school (+)	-0.65	-0.95	-0.30	(0.06)	0.32	0.14	-0.17	(0.04)
Taste shocks (+)	-0.15	-0.02	0.12	(0.06)	-1.61	-1.76	-0.14	(0.07)

Note: Predictions of two dynamic models. In a traditional model students cannot adjust study effort. In the proposed model they can because they choose the distribution of performance. Changes are with respect to the status quo prediction of each model. C-certificate: repeat grade. B-certificate = students acquired skills to proceed to next grade but only if they downgrade, i.e. switch to track of lower academic level or 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. Opportunity cost of time: $10/h. Bootstrap standard errors in parentheses.

A model where students cannot adjust their study effort leads to less favorable outcomes in both counterfactuals. For example, the increase in study delay in the "Repeat" policy is 11.6 %points instead of 9.5. The decrease in the "Downgrade" policy is 9.2 %points instead of 9.8. Also in higher education we see more negative effects if performance is exogenous. Most importantly, we would falsely conclude that there is an important negative impact on higher education graduation (-1.1 %points) from the “Downgrade” policy, while the proposed model only estimates an insignificant and small effect of -0.3.

The difference in results can be explained by an increase in study effort. 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. In the proposed model they could also change the success rate itself. Although it will be costly to do so, for many of them this might be a better option than to switch programs in advance or take the risk to fail. This extra incentive to exert study effort has important implications. Although study effort remains unobserved, we can see that students adjust it by their willingness to pay the extra cost to have better performance outcomes. This is most clear from the decrease in the number of bad performance outcomes, and especially B-certificates, in both counterfactuals. In both models the number of B-certificates goes down but the decrease in the traditional model is much smaller. In the “Repeat” policy, the decrease is 62% of the decrease in a model where students can adjust their study effort. In the “Downgrade” policy it is only 25%. This has important implications. First, there is a smaller increase in study delay in the “Repeat” policy and a stronger decrease in the “Downgrade” policy. Second, there are more students staying in more academic programs. Of course this increase in study effort comes at a cost. Both counterfactuals increase the variable costs (a component missing in a traditional model), but they do this at the benefit of other components of welfare such that the loss in total student welfare does not change significantly. The more favorable higher education outcomes compared to a traditional model are a direct result of the decrease in dropout, the increase in students graduating from academic programs, and the decrease in study delay.

We can conclude that it is important to allow students to change their effort in counterfactual simulations, not only through the choice of study program, but also through their study effort during the year. Because this remains unobserved, it can be accomplished 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. I find that 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 high school dropout. This creates large savings for society that can be reinvested in early childhood education to improve educational outcomes.

The institutional context makes it possible to do clear counterfactuals to investigate the trade-off between costs of academic effort (study effort, grade retention, risk of dropout) and the benefits in the long run (higher education degree). These conclusions are also important for other countries that track students from an early age like Germany, Austria or the Netherlands. Also 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.

From a methodological perspective, I show that it is possible to allow students to exert different amounts of study effort in counterfactual simulations, even without observing data on study effort. I also show that this is important in the application of this paper. Further research can apply the modeling strategy to other contexts. Any model where agents are expected 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. Another area of future research could relax some of the assumptions that were made. The constant marginal cost assumption on improving the odd of avoiding the lowest outcome could be problematic in counterfactuals where we expect large shifts compared to the status quo. Future research could also combine this approach with recent extensions of the traditional model along other dimensions, introducing uncertainty about the performance distribution due to imperfect information of students about their own ability (Arcidiacono et al., 2016) and endogenous quality of schools or programs due to the quality of peers and effort choices of teachers (Fu and Mehta, 2018). This would allow for counterfactuals that change the system more substantially, such as changing the age in which students are first tracked.

References

Abbring, J. H. and Daljord, Ø. (2018). Identifying the Discount Factor in Dynamic Discrete Choice Models. *Working paper.*

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.

Beffy, M., Fougere, D., and Maurel, A. (2012). Choosing the field of study in post-secondary 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. (2018a). Imperfect monitoring of job search: Structural estimation and policy design. *Journal of Labor Economics*, 36(1).

Cockx, B., Picchio, M., and Baert, S. (2018b). Modeling the effects of grade retention in high school. *Journal of Applied Econometrics*.

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

Cunha, F. and Heckman, J. J. (2009). The Economics and Psychology of Inequality and Human Development. *Journal of the European Economic Association*, 7(2/3):320–364.

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.

Fu, C. and Mehta, N. (2018). Ability Tracking, School and Parental Effort, and Student Achievement: A Structural Model and Estimation. *Journal of Labor Economics*, 36(4).

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.

Gigliotti, P. and Sorensen, L. C. (2018). Educational resources and student achievement: Evidence from the Save Harmless provision in New York State. *Economics of Education Review*, 66:167–182.

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.

Heckman, J. J. and Mosso, S. (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics*, 6(1):689–733.

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.

Jacob, B. A. and Lefgren, L. (2009). The Effect of Grade Retention on High School Completion. *American Economic Journal: Applied Economics*, 1(3):33–58.

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.

Kalouptsidi, M., Scott, P. T., and Souza-Rodrigues, E. (2016). Identification of Counterfactuals in Dynamic Discrete Choice Models.

Kapor, A., Neilson, C. A., and Zimmerman, S. (2018). Heterogeneous Beliefs and School Choice Mechanisms. *Working paper*, page 56.

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.

Lafortune, J., Rothstein, J., and Schanzenbach, D. W. (2018). School Finance Reform and the Distribution of Student Achievement. *American Economic Journal: Applied Economics*, 10(2):1–26.

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

Manacorda, M. (2012). The Cost of Grade Retention. *Review of Economics and Statistics*, 94(2):596–606.

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

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 (2012). *Education at a Glance 2012*. OECD Publishing. OCLC: 812608795.

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 Tables

Table A1: High school program and student background

Study program	Students		Male	Language ability	Math ability	High SES
All	*5158*	*(100.0%)*	*0.50*	*0.00*	*0.00*	*0.28*
Academic	*1974*	*(38.3%)*	*0.40*	*0.71*	*0.64*	*0.49*
clas+math	261	(5.1%)	0.46	1.15	1.05	0.63
clas	315	(6.1%)	0.37	0.94	0.68	0.58
math	683	(13.2%)	0.49	0.74	0.75	0.51
other	715	(13.9%)	0.32	0.41	0.36	0.38
Middle-Theoretical	*818*	(15.9%)	*0.53*	*0.11*	*0.19*	*0.22*
math	125	(2.4%)	0.70	0.32	0.47	0.30
other	693	(13.4%)	0.50	0.07	0.14	0.21
Middle-Practical	*611*	(11.8%)	*0.51*	*-0.06*	*-0.02*	*0.22*
Vocational	*1002*	*(19.4%)*	*0.51*	*-0.76*	*-0.75*	*0.10*
13th grade	609	(11.8%)	0.49	-0.67	-0.69	0.11
12th grade	393	(7.6%)	0.54	-0.89	-0.85	0.08
Dropout	*753*	*(14.6%)*	*0.67*	*-0.92*	*-0.86*	*0.07*
Part-time	431	(8.4%)	0.71	-0.97	-0.90	0.06
Full time	322	(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. dropout split between students directly opting for full time dropout or first choosing part-time option.

Table A2: High school program and higher education outcomes: summary statistics

	Higher education	
	Enrollment	Degree
Study program		
All	*58.2*	*44.0*
Academic	*96.9*	*84.2*
clas+math	99.2	94.3
clas	99.4	90.5
math	97.8	88.1
other	94.1	74.1
Middle-Theoretical	*82.0*	*51.7*
math	99.2	72.8
other	78.9	47.9
Middle-Practical	*54.8*	*27.5*
Vocational (13th grade)	13.5	*2.6*
Dropout	*0*	*0*

Note: Percentage of all students (including dropouts), 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 A3: 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*	*6.0*	*25.5*	*17.8*
Academic				
clas+math	67.0	14.2	13.0	54.0
clas	48.6	10.5	31.4	22.2
math	33.7	18.3	36.2	47.0
other	9.5	6.9	57.8	16.2
Middle-Theoretical				
math	7.2	20.8	44.8	56.0
other	0.7	3.3	43.9	17.5
Middle-Practical	*0.2*	*2.9*	*24.4*	*12.3*
Vocational (13th grade)	*0*	*0.2*	*2.5*	*0.3*

Note: Percentage of all students (including dropouts), conditional on high school program. Three types of higher education options in decreasing order of academic level: university, academic college, professional college. Graduation rates add up to the total rate of 44.0%. Each level has different programs that could be STEM. Graduation from STEM programs is reported. Clas= classical languages included. Math= intensive math.

Table A4: 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	4.4	4.4	3.8	2.5
grade 11+12	6.2	6.2	6.2	6.2	0
Middle-Theoretical					
grade 7+8	29.2	21.5	19.7	1.2	0.5
grade 9+10	100	16.5	12.4	6.4	0
grade 11+12	100	11.2	11.2	11.2	0
Middle-Practical					
grade 7+8	41.3	33.8	30.5	3.7	0.5
grade 9+10	100	100	22.3	9.3	0
grade 11+12	100	100	15.1	15.1	0
Vocational					
grade 7+8	100	100	100	7.1	0
grade 9+10	100	100	100	13.8	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 A5: Impact performance during secondary education

	Students	**High school**	**Higher education**	
		Dropout	Enrollment	Degree
All	100	14.6	58.2	44.0
At least 1 B-certificate	35.4	21.0	36.6	19.9
At least 1 C-certificate	30.0	38.6	28.9	14.7
At least 1 year of study delay	31.6	26.7	37.7	18.8

Note: First column: share of students for each performance outcome during high school. Column 2-4: share of students for each long run outcome, conditional on obtaining a bad performance outcome in high school. A-certificate: proceed to next grade, C-certificate: repeat grade, B-certificate: repeat or downgrade.

Table A6: Transitions in educational system (in % of students)

Panel A: transitions in high school

	High school outcome					
	Academic	Middle-theoretical	Middle-practical	Vocational	Dropout	Total
First track						
Academic	36.9	10.5	7.1	4.6	3.9	63.0
Middle-theoretical	1.3	4.4	3.8	6.4	3.6	19.4
Middle-practical	0.1	0.9	1.0	3.0	1.9	6.9
Vocational	0	0	0	5.5	5.3	10.8
Total	38.3	15.9	11.8	19.4	14.6	100

Panel B: transitions after high school

	Final outcome				
	Higher education degree	Higher education no degree	High school degree	High school dropout	Total
High school outcome					
Academic	32.2	5.0	1.0	0	38.3
Middle-theoretical	8.2	4.8	2.8	0	15.9
Middle-practical	3.3	3.3	5.3	0	11.8
Vocational	0.3	1.3	17.8	0	19.4
Dropout	0	0	0	14.6	14.6
Total	44.0	14.4	27.0	14.6	100.0

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

Table A7: Costs of schooling: main estimates

	Fixed costs		Log of marginal costs	
Time	1	(.)	-0.001	(0.001)
Grade	9.010	(7.804)	0.277	(0.060)
Academic				
clas+math	76.866	(53.524)	-4.845	(0.844)
clas	-141.766	(47.669)	-3.316	(0.444)
math	-12.592	(49.785)	-3.273	(0.369)
other	-225.277	(49.520)	-1.956	(0.281)
x grade	-22.407	(7.063)	0.068	(0.086)
Middle-theoretical				
math	102.125	(50.984)	-3.397	(0.465)
other	-181.509	(44.300)	-1.923	(0.245)
x grade	-7.705	(5.000)	-0.117	(0.080)
Middle-practical	-25.446	(43.907)	-1.648	(0.335)
x grade	-21.545	(4.873)	-0.218	(0.083)
Vocational	112.558	(46.284)	-4.248	(0.530)
Part-time	270.638	(31.720)		

Note: Estimates of a sample of 5,158 students or 33,239 student-year observations. Scale = minutes of daily travel time. Grade variable starts counting in high school. The marginal costs in the model are a nonparametric function of state variables, this table summarizes them by regressing their logarithmic transformation on the same variables that enter the fixed costs. Bootstrap standard errors in parentheses.

Table A8: Costs of schooling: student characteristics and elective courses

	Fixed costs				Log of marginal costs			
	Interaction with classical languages		Interaction with intensive math		Interaction with classical languages		Interaction with intensive math	
Male	-4.097	(8.294)	-50.776	(11.655)	0.527	(0.357)	0.866	(0.269)
Language ability	-57.070	(11.140)	29.557	(13.243)	-1.113	(0.410)	-0.824	(0.343)
Math ability	-21.824	(9.085)	-71.002	(14.949)	-0.479	(0.402)	0.542	(0.316)
High SES	-36.216	(9.500)	-25.388	(11.322)	-0.483	(0.391)	0.263	(0.295)
Type 2	83.543	(12.452)	35.493	(11.792)	-0.186	(0.428)	-0.022	(0.331)

Note: Estimates of a sample of 5,158 students or 33,239 student-year observations. Scale = minutes of daily travel time. The marginal costs in the model are a flexible function of state variables, this table summarizes them by regressing their 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. Clas= classical languages included. Math= intensive math. Bootstrap standard errors in parentheses.

Table A9: Type probabilities in %

	Type probabilities	
	Type 1	Type 2
Overall	*29.55*	*70.45*
Age 12	33.07	66.93
Age 13	9.86	90.14
Age 14	10.67	89.33

Note: Estimates of unobserved types in the student population by age they start high school.

Table A10: Performance thresholds

	Performance threshold for outcome	
Increase to obtain outcome 3	0.871	(0.036)
Increase to obtain outcome 4	1.102	(0.043)
Increase to obtain outcome 5	1.744	(0.054)

Note: Optimal y is specific for each grade-track and thresholds for avoiding lowest outcome in them are normalized to 0. These differences are estimated but constrained to be the same over grades and tracks. Constraints on thresholds are used to avoid impossible outcomes because of institutional context. Grade 7-10 allow more than two realizations of main performance outcome.Bootstrap standard errors in parentheses.

Table A11: Performance elective courses

	Performance	
Log variable effort ($\ln y$)		
x clas	0.902	(0.339)
x math	0.124	(0.102)
Male		
x clas	0.931	(0.469)
x math	0.058	(0.205)
Language ability		
x clas	-0.251	(0.524)
x math	0.053	(0.171)
Math ability		
x clas	-0.785	(0.532)
x math	0.488	(0.196)
SES		
x clas	-0.424	(0.352)
x math	0.027	(0.193)
Type 2		
x clas	0.332	(0.517)
x math	0.392	(0.223)
Cut points clas		
x grade	-0.096	(0.138)
x constant	2.222	(2.062)
Cut points math		
x grade 2, outcome 2	-4.679	(0.624)
x grade 2, outcome 3	-3.540	(0.556)
x grade 3, outcome 2	-5.222	(1.313)
x grade 3, outcome 3	-3.177	(0.574)
x grade 4, outcome 2	-5.676	(3.625)
x grade 4, outcome 3	-1.603	(0.528)

Note: Bootstrap standard errors in parentheses.

Table A12: Value of obtaining degree

	Degree values	
High school degree	512.195	(101.307)
x level	108.083	(60.670)
x vocational	-146.819	(99.334)
12th grade certificate vocational track	529.825	(65.461)

Note: Estimates of μ^{degree}. Scale = minutes of daily travel time. Level = academic level of high school program (0-3). Bootstrap standard errors in parentheses.

Table A13: Estimation results higher education (1)

	Higher education			
	Enrollment		Degree	
Male	-1.013	(0.108)	-0.840	(0.129)
x HE level	0.302	(0.206)	0.529	(0.246)
x STEM	0.827	(0.067)	0.598	(0.142)
Language ability	0.340	(0.126)	0.262	(0.144)
x HE level	2.130	(0.230)	0.641	(0.255)
x STEM	-0.186	(0.088)	-0.172	(0.136)
Math ability	0.111	(0.109)	0.611	(0.137)
x HE level	1.433	(0.229)	0.149	(0.335)
x STEM	0.472	(0.098)	-0.154	(0.147)
SES	0.563	(0.126)	0.633	(0.136)
x HE level	1.875	(0.191)	0.643	(0.245)
x STEM	0.084	(0.084)	-0.078	(0.129)
Type 2	-0.613	(0.157)	-1.830	(0.174)
x HE level	-4.741	(0.248)	0.901	(0.310)
x STEM	-0.639	(0.090)	1.006	(0.165)

Note: Estimates of higher education outcomes as specified in section A.4. HE Level = level of higher education (average ability). Bootstrap standard errors in parentheses.

Table A14: Estimation results higher education (2)

	Higher education			
	Enrollment		Degree	
Academic degree	4.227	(0.266)		
+ clas + math			-0.330	(0.385)
+ clas			-0.009	(0.278)
+ math			0.501	(0.198)
other			benchmark	
Middle-theoretical degree	3.426	(0.205)		
+ math			-0.195	(0.277)
other			-0.442	(0.160)
Middle-practical degree	2.050	(0.161)	-0.721	(0.233)
Vocational degree	benchmark		-2.030	(0.332)
Study delay	0.182	(0.150)	-0.580	(0.271)
High school level x study delay	-0.326	(0.077)	-0.150	(0.121)
HE level				
x high school level	0.163	(0.186)	-0.227	(0.266)
x clas	3.031	(0.296)	1.618	(0.283)
x math	2.596	(0.222)	1.313	(0.258)
x study delay	-0.312	(0.225)	0.034	(0.368)
STEM				
x high school level	-0.455	(0.062)	-0.108	(0.114)
x clas	-0.271	(0.123)	0.480	(0.193)
x math	1.335	(0.086)	0.389	(0.158)
x study delay	-0.234	(0.083)	0.165	(0.148)

Note: Estimates of higher education outcomes as specified in section A.4. Clas= classical languages included. Math= intensive math. High school level = academic level of high school program (0-3). HE Level = level of higher education (average ability). Bootstrap standard errors in parentheses.

Table A15: Estimation results higher education (3)

	Higher education			
	Enrollment		Degree	
Distance (km)	-0.018	(0.001)	-0.003	(0.001)
Same HE level as enrollment			1.715	(0.112)
Same major as enrollment			2.411	(0.088)
Upgrade HE level			-1.653	(0.283)
University	-3.957	(0.398)	-4.604	(0.447)
Academic college	-2.842	(0.339)	-2.971	(0.317)
Professional college	-1.287	(0.274)	-1.338	(0.177)
STEM	0.289	(0.171)	-1.059	(0.298)

Note: Estimates of higher education outcomes as specified in section A.4. HE Level = level of higher education (average ability). Bootstrap standard errors in parentheses.

Table A16: OLS regressions on initial conditions and counterfactuals

	Study delay (%)		**High school dropout** (%)		**Higher education graduation** (%)		**Student welfare** ($1000)	
Male	5.53	(1.41)	8.31	(1.02)	-12.40	(1.01)	-6.13	(1.67)
Language ability	-6.76	(1.24)	-7.49	(0.93)	15.38	(1.52)	16.27	(2.29)
Math ability	-1.83	(1.22)	-6.97	(0.86)	14.98	(1.25)	11.58	(1.65)
High SES	-6.46	(1.56)	-4.56	(0.74)	16.60	(1.36)	17.95	(3.02)
Type 2	11.82	(1.44)	14.36	(0.89)	-32.63	(1.32)	-38.20	(5.21)
Constant	24.01	(1.35)	1.34	(0.59)	68.53	(1.33)	70.42	(8.71)
Repeat policy	8.78	(0.73)	1.88	(0.35)	-2.04	(0.27)	-1.76	(0.25)
x male	0.85	(0.69)	1.63	(0.44)	0.05	(0.20)	-0.37	(0.12)
x language ability	-0.37	(0.63)	-1.75	(0.44)	-0.07	(0.18)	0.42	(0.17)
x math ability	-1.38	(0.69)	-0.28	(0.48)	-0.05	(0.20)	0.35	(0.16)
x high SES	-0.72	(0.74)	-1.53	(0.42)	0.16	(0.27)	0.42	(0.17)
x Type 2	0.69	(0.75)	2.38	(0.46)	0.39	(0.22)	-0.45	(0.16)
Downgrade policy	-11.91	(1.02)	-0.20	(0.22)	-0.67	(0.30)	-1.47	(0.23)
x male	-0.91	(0.70)	-1.41	(0.31)	0.30	(0.17)	-0.08	(0.12)
x language ability	-0.05	(0.68)	0.01	(0.29)	0.22	(0.13)	-0.07	(0.11)
x math ability	-0.26	(0.80)	0.95	(0.32)	-0.23	(0.14)	-0.15	(0.12)
x high SES	1.85	(0.88)	0.29	(0.32)	0.13	(0.23)	0.22	(0.17)
x Type 2	2.91	(0.97)	-1.14	(0.30)	0.26	(0.20)	0.62	(0.18)

Note: OLS regression of predictions from the dynamic model. B-certificate = students acquired skills to proceed to next grade but only if they downgrade, i.e. switch to track of lower academic level or 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. 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. Opportunity cost of time: $10/h. Bootstrap standard errors in parentheses.

Table A17: Impact conditional on B-certificate in status quo

	Study delay		**High school dropout**		**Higher education graduation**		**Student welfare**	
			Predicted value in %					
Status quo								
B-certificate in status quo	52.51	(1.38)	21.52	(1.13)	24.92	(0.88)		
No B-certificate in status quo	21.64	(0.77)	9.76	(0.54)	55.85	(1.02)		
			Change in %points				*Change in $1000*	
Repeat policy								
B-certificate in status quo	25.76	(1.28)	10.66	(0.82)	-3.58	(0.53)	-4.47	(0.56)
No B-certificate in status quo	-0.30	(0.14)	-0.10	(0.07)	-0.57	(0.09)	-0.73	(0.09)
Downgrade policy								
B-certificate in status quo	-24.80	(1.18)	-4.20	(0.62)	-0.46	(0.43)	-1.56	(0.23)
No B-certificate in status quo	-0.83	(0.12)	-0.05	(0.04)	-0.21	(0.06)	-0.69	(0.09)

Note: Predictions conditional on obtaining B-certificate in status quo. B-certificate = students acquired skills to proceed to next grade but only if they downgrade, i.e. switch to track of lower academic level or 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. Opportunity cost of time: $10/h. Bootstrap standard errors in parentheses.

A.2 Data appendix

A.2.1 The LOSO dataset

The dataset used for this paper is the LOSO dataset.[25] The first part of the data contains rich information of students and their parents, and choices and performance measures during high school in the region of Flanders (Belgium). We can follow a cohort of students starting high school in 1990. I also include results from follow-up research, called "LOSO-annex", which looked into the education and labor market career in the first three years after leaving high school (academic years starting in 1996 until 1998 for most students, but later for those with study delay). This data was later enriched by sending questionnaires during 2003-2005 to students that were still in the educational system in the questionnaire before.

The students are not randomly selected over Flanders as a whole but instead two large subregions of Flanders were defined that are considered to be representative for the entire region.[26] In these regions almost all schools are included, and within each school every student is included. The first subregion is in the east part of Flanders and includes the municipalities Hasselt, Genk, Beringen, Leopoldsburg, Herk de Stad and Diest. The second subregion is more to the west and contains the schools in Dendermonde, Hamme and Zele. Data was collected from students, parents, teachers and schools and they were actively contacted by researchers at multiple occasions. This is why the data is of high quality and there is very few attrition. Even if a student decides to leave his school for a school that was not initially part of the project, it was still possible to collect the necessary information.

[25]See also https://ppw.kuleuven.be/o_en_o/COE/losodatabank.

[26]To test the representativeness of the data, I compared higher education enrollment number (58%) to population data. For Belgium as a whole, I find an almost identical number around the same time period: 56% in 1996 and 57% in 1999 (UNESCO Institute for Statistics, indicator SE.TER.ENRR).

A.2.2 Sample selection

I only keep the 6,439 students in the dataset that are known as 'proefgroepleerlingen'. These are students that are tracked from the start of high school, even if they move to another school. The dataset also contains a large number of observations of inflow in schools over time but these are not used in this study. From these students, I eventually keep 5,158 students to estimate the model.

The model in this paper captures the main aspects of the education system, but also makes some simplifications such that it cannot explain every observation in the data. Moreover, some data on choices or outcomes that is needed for the estimation is missing. Table A18 summarizes the attrition. More details on why observations had to be dropped follow next.

A.2.3 Data interpretation

Some information in the data is not straightforward to use in the model. Therefore, I create or adjust some of the information to capture the spirit of the educational system with the model, without overly complicating it to capture all anomalies in the data. In particular, I perform the following manipulations.

First, students who are successful in the first grade of the vocational track have the possibility to go to the first grade of another track. I do not allow for this possibility in the model and instead make these students look as if they entered the non-vocational track after an additional year of study delay in elementary school. Second, B-certificates often exclude specific programs like technical education-science, or accountancy-informatics, and not always entire study programs as defined in the model. In many cases, only "unrealistic" alternatives remain within the same study program that I include in the model (e.g. a program that is not available in any school in the neighborhood). To avoid modeling every single study program, as well as school choice, I instead use a model with aggregated study programs and interpret the certificate data in the following way.

Table A18: Data attrition

	Total number of students	**Loss**	**Relative loss compared to start**	**Reason for dropping**
Data description	6439			
Data received (includes birth data and gender)	6411	-28	0.004	missing data
No students that leave and return to secondary education	6381	-30	0.009	not allowed by model
No time period	6365	-16	0.011	missing data
No performance	6327	-38	0.017	missing data
No study program	6302	-25	0.021	missing data
Do not allow switch from middle to vocational after grade 11	6263	-39	0.027	not allowed by model
Do not allow to skip grades in high school	6260	-3	0.028	not allowed by model
Go down grades in high school	6249	-11	0.030	not allowed by model
Students have to start in the first grade of high school	6246	-3	0.030	not allowed by model
Ignore students that skipped grade in elementary school	6212	-34	0.035	not allowed by model
Students make choices that are inconsistent with the certificate they received and/or track they were in	5936	-276	0.078	not allowed by model
Students that move from part time to full time education	5903	-33	0.083	not allowed by model
Students that drop out illegally	5832	-71	0.094	not allowed by model
No info on choice after leaving high school	5705	-127	0.114	missing data
No info on obtaining a higher education degree within 6 years	5558	-147	0.137	missing data
No info on location of student	5442	-116	0.155	missing data
Missing characteristics or test score of students	5179	-263	0.196	missing data
Do not allow 3 or more years of study delay at start high school	5177	-2	0.196	not allowed by model
Do not allow students to live more than 50 km to closest school of each type	5160	-17	0.199	remove outliers
Do not allow to go to higher education without high school degree	5158	-2	0.199	not allowed by model

Certificates that exclude an entire track are straightforward to implement. This already contains 67% of the data on B-certificates. In other cases, I proceed as follows. I always assume a hierarchy: if a low track is excluded, the higher ones are excluded too.[27] I also use a slightly different definition of a B-certificate that is more consistent over the different grades. I ignore the officially called "C-certificates" in grade 7 as they do not restrict entry into grade 8 of the vocational track, and change them to B-certificates that allow the vocational track in the next grade (or A-certificate if the student is already in the vocational track). In other cases in the academic and middle-theoretical track, I use the following procedure. This procedure was established to be in line as much as possible with the spirit of the educational system, as well as to minimize the number of choices in the data that would not be possible to be explained by the model. I make groups of aggregated study programs that are less aggregated than the ones used in the model, but more aggregated than how they appear in the data. This aggregates over very small differences within programs between which a B-certificate is not expected to ever make a distinction, except when teachers (and probably students) are not aware of the existence of the program. A B-certificate then excludes all classical language options if all the aggregated programs with classical languages appear in the list of restrictions. It excludes math options and the entire track if there is an exclusion within all the major aggregated options of these study programs. For exclusion of the middle-practical track, one occurrence of a program in the track in the list of restrictions restricts the entire track, unless choice behavior and the corresponding grade is not consistent with that.

At this point we went from explaining 67% of the B-certificate data to explaining 95%. The remaining 5% is assumed to be imposing irrelevant restrictions to the students in the model and are replaced by A-certificates. An important part of this 5% also contains exclusions within the vocational track which are unrelated to the

[27]The following example shows that this is reasonable to assume: out of 199 B-certificates that exclude all programs in the middle tracks for students currently in an academic track, 197 certificates also exclude the academic track.

academic level of the program and are therefore outside the scope of this paper.

A.2.4 Details about study programs

The official distinction between tracks differs slightly from the one proposed in the paper. The official track names are "ASO", "TSO", "KSO", "BSO" and "BUSO" and the distinction for most tracks is made from the third year on (i.e. grade 9). ASO corresponds to the academic track, BSO and BUSO to the vocational track and both TSO and KSO are middle tracks (that differ in their focus on respectively technical education and artistic education). I then split up this middle track according to programs that prepare primarily for higher education (middle-theoretical) and the labor market (middle-practical), which is a common distinction made, e.g. in Cockx *et al.* (2018b), but also by the researchers that collected the data.[28]

Although this official distinction does not exist in the first two grades of high school, there is a distinction between programs preparing for the different tracks. First of all, there is the distinction between a B-stream, preparing for the vocational track only, and an A-stream, preparing for the other tracks. Within the A-stream one can also distinguish between more or less theoretical programs, based on the hours per week each school can decide what to teach (5 in grade 7 and up to 10 in grade 8). This distinction was made by the LOSO researchers, although not directly linked to the specific track they prepare for. Therefore, I looked at the most common transition patterns to assign them to a track. In a few cases the distinction within the A-stream was not made, I then assumed students were in the same track as the year after.

As mentioned in Cockx *et al.* (2018b), upward mobility is theoretically possible but practically infeasible which is why it almost never occurs in the data. Nevertheless, I do allow for this flexibility in non-vocational tracks in the first two grades as I

[28]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 do 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.

do see some upward mobility when the official track structure is not yet established. Note that any mobility between grade 11 and grade 12 is forbidden, except for a switch between some programs from a middle track to the vocational track. I do not allow for that in the model and drop the students that do this. I also exclude the following uncommon choices in the model: dropping out of (full time) high school and returning, and repeating the grade in a track of higher academic level or with an elective course that was not chosen before. Furthermore, sometimes rules are not strictly followed. Some cases can be illegal, but in other cases parents could have asked for special permissions from teachers, the ministry of education or as a result from a court order. These special cases are dropped.

For the higher education options, the distinction between different levels is also used in official statistics on Flemish education and corresponds to respectively “Hoger onderwijs van het korte type”, “Hoger onderwijs van het lange type” and “Universiteit”. Today, the distinction between “Hoger onderwijs van het lange type” and “Universiteit” is no longer made but the specific programs within them are still similar. To define STEM majors, I also use a characterization by the Flemish government (https://www.onderwijskiezer.be/). The different types of (higher) education are associated with large differences in wages. To demonstrate this, I use data of the "Vacature Salarisenquête", a large survey of workers in Flanders in 2006, and compare the median wages of 30-39 year olds (sample size of 20,534 workers). High school dropouts earned a gross monthly wage of 2,039 EUR, high school graduates without a higher education degree earned 2,250 EUR, professional college graduates 2,600 EUR, academic college graduates 3,281 EUR and university graduates 3,490 EUR. Students that graduated in a STEM major earned 3,264 EUR, while students that graduated in a non-STEM major earned 2,800 EUR.

A.2.5 Distance and travel time data

I use address data of students and schools to obtain coordinates using the Stata command "geocode3". For the schools I updated this manually when geocode returned an error or was not very precise. I did this for schools with at least 10 student-time observations using Google maps. I then use the "osrmtime" command to calculate travel time by bike to the closest school that offers the study program.[29] Note that all schools attended by students in the sample are used, which includes also schools outside of the ones assigned by the researchers (because students can switch to other schools). I dropped students living more than 50km from any school as they are more likely to be influenced by schools that I do not observe or are outliers because of measurement error when geocoding.

At the higher education level I look at the distance to the closest school for each option (level and major) if it is not a university and I distinguish between the five Flemish campuses for universities (Leuven, Ghent, Brussels, Antwerp and Diepenbeek). This is similar to Declercq and Verboven (2018). If students attend a university abroad or in Wallonia, I assign them randomly to one of the Flemish campuses, using a probability distribution that corresponds to the distribution of students going to Flemish universities.

A.2.6 Policy relevance

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. Therefore, it is 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

[29] A bike is the most popular mode of transportation. 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.

students have the right incentives within the system to optimize total welfare (OECD, 2017). 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/student or 11% of total expenditures on compulsory education, the highest rate in the OECD (OECD, 2013).

A.3 Curvature of marginal benefits

This section discusses the shape of the marginal benefits of effort that were introduced in section 3. I show that the marginal benefits are positive, decreasing in effort and follow an S-shaped curve.[30]

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 $P_j(g_{it+1} = 1|y_{it}) = P_j(g_{it+1} \leq 1|y_{it})$. Therefore, we can write the marginal benefits as follows:

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

Because performance shocks are distributed logistically, we know that $\frac{\partial P_j(g_{it+1} \leq \bar{g}|y_{it})}{\partial y_{it}} = -\frac{(1-P_j(g_{it+1} \leq \bar{g}|y_{it}))P_j(g_{it+1} \leq \bar{g}|y_{it})}{y_{it}}$. 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 in-

[30]Note that in the appendix I explain how I extend the model to allow for additional performance measures. In rare cases this can lead the marginal benefits to be increasing such that the first-order condition can have more than one solution. Nevertheless, estimation can still proceed as explained. This is because we only need the necessary condition for an optimum to identify the marginal costs. It is however important to do a grid search in counterfactuals to look for the global optimum.

tuitive: the marginal benefit is larger with large gains of getting a higher performance outcome, but less so if the variable effort component is already high.

Marginal benefits are decreasing in effort

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

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

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

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 marginal benefits 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.4 Estimation details

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 discuss the CCP representation and I finish by discussing the likelihood function.

A.4.1 Choice set

Each study program in high school ($j \in se$) 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.[31] 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.[32] As explained in the institutional context, the flexibility of a B-certificate can have different implications for the choice set. Therefore, I 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 here to allow for

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

[32] There is one exception. Students can enroll in grade 8 of the vocational track without having succeeded grade 7. Therefore, the lowest performance outcome is a B-certificate and students in grade 7 of the vocational track do not pay a variable cost.

additional performance measures that contain permissions to study classical languages g_{it+1}^{clas} and intensive math g_{it+1}^{math}.

From the age of 18 on, students have the possibility to leave the education system: $j \in (0, he)$ with $j \in he$ enrollment in a higher education (only in the choice set if the student 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.4.2 Fixed costs

In this section I discuss functional form assumptions on fixed cost.[33] Let C_{ijt}^0 be the fixed cost of student i in option j at time t:

$$\begin{aligned}
C_{ijt}^0 = {} & \mu_j^0 + \mu_j^{\text{grade}} \text{grade}_{ijt} \\
& + S_i'(\mu_j^{S,0} + \mu^{S,\text{level}} \text{level_SE}_{ijt} + \mu^{S,\text{math}} \text{math}_{ijt} + \mu^{S,\text{clas}} \text{clas}_{ijt}) \\
& + \nu_i'(\mu_j^{\nu,0} + \mu^{\nu,\text{level}} \text{level_SE}_{ijt} + \mu^{\nu,\text{math}} \text{math}_{ijt} + \mu^{\nu,\text{clas}} \text{clas}_{ijt}) \\
& + \mu_{time} \text{time}_{ijt} \\
& + \text{retention}_{ijt}'(\mu^{\text{ret},0} + \mu^{\text{ret,level}} \text{level_SE}_{ijt}) \\
& + \mu_{up} \text{upgrade}_{ijt} + \mu_{down} \text{downgrade}_{ijt} \\
& + \mu_{staymath} \text{math}_{ijt} \times \text{math}_{it-1} + \mu_{stayclas} \text{clas}_{ijt} \times \text{clas}_{it-1}.
\end{aligned}$$

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 to the closest school that offers the study option in the current grade and grade_{ijt} is the grade a student is in (set such that 1 is the first year of high school). Level_SE_{ijt} is the academic level of the track a student 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 clas refer to

[33] Note that the part-time track does not have a grade structure. Therefore, I 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 dropout.

respectively programs with intensive math and with classical languages. Grade retention 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 ("Repeat") and a stock variable that captures the years of study delay accumulated in previous years ("Study delay"). 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 and $\mu_{staymath}$ and $\mu_{stayclas}$ capture preferences to stay in a program with the same elective courses.

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 rescale the parameters by dividing by μ_{time}. This way, the cost estimates can be measured in daily commuting time.

A.4.3 Performance

Main performance outcome

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 $g_{it+1} = \{1, 2, ..., 5\}$. If $g_{it+1} = 1$, student i's performance at time t was insufficient to go to the next grade, regardless of the program they want to follow. $g_{it+1} = 2$ allows access to the next grade of the vocational track (*voc*) but not other tracks. Similarly, $g_{it+1} = 3$ additionally allows access to the next grade *midp*, $g_{it+1} = 4$ allows *midt* and $g_{it+1} = 5$ 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. Therefore, I approximate the y^*_{ijt} by a (flexible) parametric structure. The optimal levels of effort and the thresholds to obtain each outcome can then be recovered by estimating an ordered logit model with index $\ln(y^*_{ijt})$. The functional form of the index is similar to what is imposed for the fixed cost parameters, using a linear in parameters specification for each observed and unobserved student characteristic, but I allow for more flexibility by letting them be track-specific and change linearly over different grades. I also allow distance to higher education options to affect performance and I add an effect of lagged study program (academic level and dummy variables for intensive math and classical languages). 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 in. Because there is little variation in the data, I only estimate three parameters after the normalization to capture the increase in thresholds for obtaining a higher outcome, which is then assumed to be constant over grades and tracks (see Table A10).

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. Therefore, I extend the model to allow for two additional measures of performance: $g^{clas}_{it+1} = \{1, 2\}$ specifies if a students can go to the next grade in a program that includes classical languages. $g^{math}_{it+1} = \{1, 2, 3\}$ specifies if a student can go to the next grade in a math option in the middle-theoretical track ($g^{math}_{it+1} = 2$) or the academic track ($g^{math}_{it+1} = 3$).[34] I model their distribution by an ordered logit, conditional on the

[34]An implication of this extension is that marginal benefits of effort in the model are not necessarily decreasing over the entire domain of y, 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.

outcome of g_{it+1}, with indexes:

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

I also estimate grade- and track-specific thresholds.[35] $\alpha_y^{math} > 0$ and $\alpha_y^{clas} > 0$ measure how much of the variable effort component, identified from the permissions to start in each track, matters for each elective course. I also allow for comparative advantage in elective courses by estimating the influence of observed and unobserved student characteristics through $(\alpha_S^{math}, \alpha_\nu^{math})$ and $(\alpha_S^{clas}, \alpha_\nu^{clas})$. The results can be found in Table A11.

A.4.4 Higher education

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 hereby impose structure on the choice-specific

[35]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 $g_{it+1}^{track} = 4$, $g_{it+1}^{math} \in \{1, 2\}$ and $g_{it+1}^{clas} = 1$.

component $\Psi_j^{HEE}(x_{it}, \nu_i) = \Psi_{ij}^{HEE}$:

$$\begin{aligned}
\Psi_{ij}^{HEE} &= \phi_j^{HEE,0} && (16)\\
&+ S_i'(\phi^{HEE,S,0} + \phi^{HEE,S,level}\text{level_HE}_j + \phi^{HEE,S,STEM}\text{STEM}_j)\\
&+ \nu_i'(\phi^{HEE,\nu,0} + \phi^{HEE,\nu,level}\text{level_HE}_j + \phi^{HEE,\nu,STEM}\text{STEM}_j)\\
&+ \phi^{HEE,\text{dist}}\text{distance_HE}_{ij}\\
&+ d'_{iT_i^{SE}}\phi^{HEE,\text{SE}}\\
&+ \text{delay}_{iT_i^{SE}}(\phi^{HEE,delay,0} + \phi^{HEE,delay,level}\text{level_HE}_j + \phi^{HEE,delay,STEM}\text{STEM}_j)\\
&+ \phi^{HEE,levelxdelay}\text{level_SE}_{iT_i^{SE}} \times \text{delay}_{iT_i^{SE}} && (17)\\
&+ X_{ij}'\phi^{HEE,\text{interact}}
\end{aligned}$$

Level_HE$_j$ is the level of the higher education program. I follow Arcidiacono (2005) and define the level for each type of higher education by the average math ability of the enrolling students. I use professional college as a benchmark (0.20) and calculate differences with academic college (0.59) and university (0.79). 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 from and delay$_{iT_i^{SE}}$ the years of accumulated study delay. Since there are few students in the academic track that do not enroll in higher education, I do not distinguish between elective courses and estimate a common effect of each track on enrollment in the benchmark professional college. I also include a vector of interactions X_{ij} 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).

I impose a similar model for graduation from higher education. I use a similar functional form as in (16) but I also add interaction effects in X to take into account the enrollment decision.[36] In particular I include dummy variables for choosing the

[36]In contrast to college enrollment rates, there is sufficient variation in graduation rates within

same level, upgrading a level, and choosing the same major. I add a 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 decisions and the final degree a student obtains.

A.4.5 CCP representation 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 leave secondary education in the outside option (no higher education). The exposition in this section is similar to Arcidiacono and Miller (2011) and Arcidiacono and Ellickson (2011).

The choice probabilities (9) at the optimal levels of the variable effort component can be written by using differenced value functions:

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

programs of the same track. Therefore, I do not need to restrict the common parameters of the effect of study programs to be the same.

$$\begin{aligned}
&\text{with } v_{ijt}(x_{it},\nu_i,y^*_{ijt}) - v_{ij't}(x_{it},\nu_i,y^*_{ij't}) \qquad (18)\\
&= u_j(x_{it},\nu_i) + \beta\sum_{\bar{g}} P_j(g_{it+1}=\bar{g}|y^*_{ijt})\bar{V}_{t+1}(x_{it+1}(\bar{g}))\\
&- u_{j'}(x_{it},\nu_i) - \beta\sum_{\bar{g}} P_{j'}(g_{it+1}=\bar{g}|y^*_{ij't})\bar{V}_{t+1}(x_{it+1}(\bar{g})),
\end{aligned}$$

for any $j^{'} \in \Phi(x_{it})$ and $u_j(x_{it},\nu_i) = -C^0_j(x_{it},\nu_i) - c_j(x_{it},\nu_i)y^*_{ijt}$, 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 (13) in (18):

$$\begin{aligned}
&v_{ijt}(x_{it},\nu_i,y^*_{ijt}) - v_{ij't}(x_{it},\nu_i,y^*_{ij't}) \qquad (19)\\
&= u_j(x_{it},\nu_i) + \beta\sum_{\bar{g}} P_j(g_{it+1}=\bar{g}|y^*_{ijt})\left(\gamma + v_{id^*t+1}(x_{it+1}(\bar{g}),\nu_i) - \ln\Pr(d^*_{it+1}|x_{it+1}(\bar{g}),\nu_i)\right)\\
&- u_{j'}(x_{it},\nu_i) - \beta\sum_{\bar{g}} P_{j'}(g_{it+1}=\bar{g}|y^*_{ij't})\left(\gamma + v_{id^*t+1}(x_{it+1}(\bar{g}),\nu_i) - \ln\Pr(d^*_{it+1}|x_{it+1}(\bar{g}),\nu_i)\right)
\end{aligned}$$

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 $d^*_i = (d_{it}, d^*_{it+1}, d^*_{it+2}, ...d^*_{it+\rho_{it}})$:

$$\begin{aligned}
\kappa^*_\tau(g_{i\tau+1}|x_{it},\nu_i) &= \sum_{\bar{g}} P_{d^*}(g_{i\tau+1}=\bar{g}|y^*_{d^*t}(x_{i\tau},\nu_i)) \text{ if } \tau = t \qquad (20)\\
\kappa^*_\tau(g_{i\tau+1}|x_{it},\nu_i) &= \sum_{\bar{g}} P_{d^*}(g_{i\tau+1}=\bar{g}|y^*_{d^*t}(x_{i\tau},\nu_i))\kappa^*_{\tau-1}(g_{i\tau}|x_{it},\nu_i) \text{ if } \tau > t
\end{aligned}$$

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 $\kappa^{'}_\tau$ to be the transitions in a sequence where the choice in t is different: $d^{'}_i = (d'_{it}, d^*_{it+1}, d^*_{it+2}, ...d^*_{it+\rho_{it}})$.[37] We can then repeat the CCP method

[37]We can also allow a more general alternative sequence in which the choice in each period is

in each of the future periods and rewrite (19) as the sum of future flow utilities and CCPs until the outside option becomes available at $t + \rho_{it}$:

$$
\begin{aligned}
&v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) - v_{ij't}(x_{it}, \nu_i, y^*_{ij't}) \qquad (21)\\
&= u_j(x_{it}, \nu_i) - u_{j'}(x_{it}, \nu_i)\\
&+ \beta[u_{d^*}(x_{it+1}(g_{it+1}), \nu_i) - \ln \Pr(d^*_{it+1}|x_{it+1}(g_{it+1}), \nu_i)]\kappa^*_t(g_{it+1}|x_{it}, \nu_i)\\
&- \beta[u_{d^*}(x_{it+1}(g_{it+1}), \nu_i) - \ln \Pr(d^*_{it+1}|x_{it+1}(g_{it+1}), \nu_i)]\kappa'_t(g_{it+1}|x_{it}, \nu_i)\\
&+ \sum_{\tau=t+2}^{t+\rho_{it}-1} \beta^{\tau-t}[u_{d^*}(x_{i\tau}(g_{i\tau}), \nu_i) - \ln \Pr(d^*_{i\tau}|x_{i\tau}(g_{i\tau}), \nu_i)]\kappa^*_{\tau-1}(g_{i\tau}|x_{it}, \nu_i)\\
&- \sum_{\tau=t+2}^{t+\rho_{it}-1} \beta^{\tau-t}[u_{d^*}(x_{i\tau}(g_{i\tau}), \nu_i) - \ln \Pr(d^*_{i\tau}|x_{i\tau}(g_{i\tau}), \nu_i)]\kappa'_{\tau-1}(g_{i\tau}|x_{it}, \nu_i)\\
&+ \beta^{\rho_{it}}\overline{V}_{t+\rho_{it}}(x_{t+\rho_{it}}(g_{it+\rho_{it}}), \nu_i)\kappa^*_{t+\rho_{it}-1}(g_{it+\rho_{it}}|x_{it}, \nu_i)\\
&- \beta^{\rho_{it}}\overline{V}_{t+\rho_{it}}(x_{t+\rho_{it}}(g_{it+\rho_{it}}), \nu_i)\kappa'_{t+\rho_{it}-1}(g_{it+\rho_{it}}|x_{it}, \nu_i).
\end{aligned}
$$

$\overline{V}_{t+\rho_{it}}$, the value of behaving optimally when the outside option is available, can be written as in (14). 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.

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

different but here it is sufficient to only let the first choice be different.

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 replacing the marginal cost of effort by the marginal benefit of effort in the data, future value terms also enter directly into $u_j(x_{it}, \nu_i)$ (see (12)). Because $\sum_{\bar{g}} \frac{\partial P_j(\bar{g}|y_{it})}{\partial y_{it}} = 0$, all terms that do not depend on performance drop out such that the same simplifications arise because of finite dependence.

A.4.6 Approximating the CCPs

Ideally, the CCPs are recovered for each realization of the state before estimating the structural parameters of the model. However, the large number of states makes this impossible. Therefore, I follow Arcidiacono *et al.* (2016) and use predictions of a flexible conditional logit to approximate them. Similar to the index that predicts the log of the variable effort component in the data, I assume a functional form that is linear in observed and unobserved characteristics for each student characteristic, and I allow for more flexibility than in fixed costs by letting them be track-specific and change linearly over different grades. I also allow distance to higher education options to affect choices, while they are excluded from fixed costs.

A.4.7 Likelihood function to estimate

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.[38] Let the fixed cost parameters in $C_j^0(.)$ and the common value of a degree μ^{degree} be given by μ and the parameters that predict higher education outcomes by $\phi = (\phi^{HEE}, \phi^{HED})$. The loglikelihood function is then

$$\ln L_i(\mu, \phi) = \ln \sum_{m=1}^{M} \pi_{m|age\_start} L_i^m(\mu, \phi)$$

with $L_i^m(\mu, \phi) = \prod_{t=1}^{T_i^{SE}} L_{it}^{program,m}(\mu, \phi^{HEE}) \times L_{it+1}^{performance,m} \times L_{it}^{ccp,m} \times L_i^{HEE,m}(\phi^{HEE}) \times L_i^{HED,m}(\phi^{HED})$

with $L_{it}^{program,m}$ and $L_i^{HEE,m}$ given by logit choice probabilities (9), with conditional value functions (15) and (6). $L_i^{HED,m}(\phi^{HED})$ is given by the conditional logit probabilities on the different possibilities for higher education graduation outcomes. The likelihood contribution of the performance outcome in secondary education is given by ordered logit probabilities $L_{it+1}^{\text{performance},m}$ and $L_{it}^{ccp,m}$ are the CCP predictors. Note that the inclusion of unobserved types makes the function no longer additively separable such that sequential estimation is not possible anymore.

Arcidiacono and Miller (2011) show that additive separability can be restored.

[38]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.

The estimation procedure is an adaptation of the EM algorithm. It starts from a random probability of each observation to belong to each type. The entire model can then be estimated as explained in the main text, but weighs each observation-type combination by the probability that the student 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 $\bar{\eta}_{j,grade}$ are identified in a first stage.[39] 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 (8) is used to recover the marginal costs.

A.5 Details about simulations

All predicted values are calculated as follows. I first categorize students by their demographic characteristics: gender, language ability, math ability, SES and the age they start high school. I discretize the observed ability distribution by creating four equally sized groups for each measure. Every student then belongs to one group which is a unique combination of these variables. Within each group I use the average travel times and distances. Each group is then used to calculate the value functions for each unobserved type. To limit the number of calculations, I drop groups with less than

[39] 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 effort costs. Also for counterfactual analyses, the full model is needed.

10 students and verify that this has a negligible effect on the distribution of student characteristics.

After obtaining the value functions, I proceed to simulation during high school. For each type I draw 10,000 students using the empirical distribution of the observable characteristics. I also take draws of taste shocks for every option in every period, as well as performance shocks in every period for every performance outcome. The average statistics are then calculated on a total of 20,000 draws. Given the simulated outcomes of high school, I used the closed form expressions for higher education to calculate enrollment and graduation.

This procedure allows for a substantial total number of draws, while needing only a limited number of students to use for a grid search to find the optimal effort level within each possible program. The grid search for effort levels starts at the optimal value of the scalar y in the data and looks for better levels using five sequential loops and an additional step to check for a corner solution. The first loop looks at changes in the log of the variable effort component by 1 unit with a minimum of -5 and a maximum of +5. The second loop divides steps and thresholds by five, the third by 5, the fourth by 25 and the fifth by 625, such that the final precision is 0.0016 (which is about 0.16% for the variable effort component y). Finally, I check if a corner solution is optimal by setting $y = 0$ and changing the performance distribution to predict the worse outcome with probability 1.

Standard errors are obtained by using the different estimates of each bootstrap sample and by repeating the entire procedure for each of them.

A.5.1 Predicting higher education outcomes

Note that to evaluate the impact on higher education outcomes, a structural model in high school 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 same primitives of the model but only the way these outcomes are influenced by secondary education outcomes, after controlling for observed and unobserved student characteristics. Therefore, I 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 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 in the status quo scenario.

Table A19 shows the ability of the model to replicate the actual data. The model does a good job in predicting the patterns in the data such that it can be used for counterfactual simulations. We see that graduation rates in different track and higher education outcomes are predicted very precisely. There is a slight overprediction in the number of students with a B-certificate leading to a small overprediction in the number of students with study delay.

Table A19: Predictions of the model

		Data	Predictions	
High school (% of students)				
Academic		38.27	40.02	(2.07)
	clas+math	5.06	5.03	(0.65)
	clas	6.11	3.18	(0.42)
	math	13.24	14.59	(1.27)
	other	13.86	17.22	(1.29)
Middle-Theoretical		15.86	16.10	(1.24)
	math	2.42	3.11	(0.46)
	other	13.44	12.99	(0.97)
Middle-Practical		11.85	8.14	(1.19)
Vocational		19.43	21.57	(0.89)
Dropout		14.60	14.17	(0.67)
Students with at least 1 B-certificate		35.40	37.53	(0.81)
Students with at least 1 C-certificate		30.01	30.69	(0.77)
Students with at least 1 year of study delay		31.62	33.22	(0.91)
Higher education (% of students)				
Enrollment		58.18	58.15	(0.75)
Graduation		44.01	44.25	(0.75)
University degree		12.43	11.22	(0.55)
Academic college degree		6.05	6.26	(0.38)
Professional college degree		25.53	26.77	(0.69)
Degree in STEM major		17.76	18.01	(0.65)

Note: Clas= classical languages included. Math= intensive math. Observed outcomes in the data and predictions from the proposed dynamic model. Bootstrap standard errors of predicted values in parentheses.

A.5.2 Welfare analysis

Opportunity cost

I assume an opportunity cost of \$10/hour. This is chosen to approximate the opportunity cost of students in high school and is consistent with Kapor et al. (2018). Students are not allowed to work until they are 15 years old and the wage often depends on their age. In 2012 the minimum wage ranged between €6.8 and €9.7/hour (https://www.jobat.be/nl/artikels/wat-is-het-minimumloon-voor-een-jobstudent/). Only a small amount of taxes is paid on this if they work a limited amount of hours. To compare to OECD estimates, I use the PPP adjusted exchange rate of dollars (0.82), which results in wages between \$8 and \$12. Note that the model is in years while the estimates are scaled in minutes/day. Therefore, I multiply them by the wage per minute (\$10/60) and the 177 school days there are in a year.

Gains from reducing grade retention

The direct cost and the total foregone earnings can be found in Table IV.1.6 in OECD (2013). I subtract the net income (49%) to only capture the externality. This number was calculated by dividing column (7) by column (1) in Table A10.2 in OECD (2012).

Reinvestment of gains

Recent estimates for the effect of a one-time "helicopter drop" increase of \$1,000 on the ability distribution are around 1% to 2% of a standard deviation (Gigliotti and Sorensen, 2018; Lafortune et al., 2018). For this reason, reinvesting the efficiency gains of the "Downgrade" policy could result in substantial gains for students. Using the estimates in Table A16 and the savings from avoiding grade retention ($\$2,910$), a 1.5% effect per \$1,000 on each of the observed ability measures in the downgrade policy would bring back $\$1,200$ in student welfare, increase graduation rates in higher education by 1.3 %points, reduce study delay by 0.4 %points and dropout by 0.6 %points. This in turn also creates additional savings that could be reinvested.

The estimates should be interpreted with caution. First, gender, socioeconomic status and the unobserved type are capturing all initial skills that are not captured by the language and math ability measures. A policy that changes skills might therefore have a bigger effect than estimated now. On the other hand, ability measures could also capture other things that might not responds to increased funding, e.g. parental characteristics that are not captured by the SES dummy.