The effects of high school curriculum. A model of program and effort choice.

Olivier De Groote*

November 24, 2017

Abstract

This paper addresses the impact of study programs in secondary education on long run educational and labor market outcomes. I estimate a dynamic model of educational decisions that allows for observed and unobserved differences in initial ability. It is novel in that it adds unobserved effort as a choice variable, along with the choice of study program. This replaces traditional approaches, which assume end-of-year performance follows an exogenous law of motion. I use the model to calculate how each study program contributes to different outcomes and I investigate policies that aim to match students to the right program. I find that academically rigorous programs are important to improve higher education outcomes, while vocational programs prevent drop out, grade retention and unemployment. At the same time, policies that encourage underperforming students to switch to less academic programs do not have a negative impact on higher education outcomes and they substantially reduce grade retention and drop out. I also find that ignoring the fact that students choose their effort level generates biases in counterfactual predictions.

*KU Leuven and Ph.D. fellow of the Research Foundation Flanders (FWO), email: olivier.degroote@kuleuven.be. This paper benefited from helpful comments at various stages from Peter Arcidiacono, Jan De Loecker, Koen Declercq, Arnaud Maurel, Jo Van Biesebroeck and Frank Verboven, several audiences at KU Leuven and Duke University and participants of the IWAEE 2017 and EALE 2017 conferences. 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 Flemish 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 programs in higher education. To achieve the European 2020 target of 40% college educated people, many countries aim to induce more students to choose academic curricula. Other countries often provide course-level differentiation. In the US, there is a similar trend towards more academic course taking, especially in STEM (Science, Technology, Engineering, Math)-fields.[1]

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

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

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

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

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

retention, at the cost of graduation rates of academic programs.

I develop a dynamic model of educational decisions. In each year of secondary education students choose a program based on the (psychic) cost of studying today and the impact on future utility. The cost consist of (1) a fixed cost of following the program and (2) a variable cost, increasing in unobserved study effort. I distinguish between both costs by modeling end-of-year performance as a function of the chosen level of study effort. Students can change effort by choosing the distribution of performance outcomes. A better distribution is costly today, but leads (in expectation) to better opportunities and outcomes in the future. Identification of the variable costs then follows from a first-order condition that sets the (unobserved) marginal cost of effort today equal to the (observed) expected marginal benefit of effort in the future. After accounting for the variable costs and the impact on the future, fixed costs can be identified from the remaining variation in the data on program choices. In the application, performance takes the form of permissions to different study programs in the next grade during secondary education, and obtaining a degree after the last grade. Allowing for dynamics in the choice of both program and study effort is therefore particularly important as we expect students to adapt their current program and level of study effort to changes in tracking policies.

To look at the long-run impact of study programs, I simultaneously estimate their effect on outcomes in higher education (enrollment, first year performance and graduation at age 25) and the unemployment spell between age 25 and age 35. To identify causal effects, it is crucial to allow for correlation between the unobservables that impact long run outcomes and choices in secondary education, e.g. because of unobserved ability. I model this using a finite mixture of types with each type affecting all components of the model. Rich panel data and exclusion restrictions help in identifying the types without relying on arbitrary functional form assumptions. The Flemish context is particularly useful for this purpose as students make repeated choices and obtain important performance outcomes during many years. The extensive data on initial ability and socioeconomic status also allows me to include rich patterns of heterogeneity, even with a small number of unobserved types.

This paper contributes to three strands of literature. A first strand of literature investigates the causal impact of high school curriculum on long run educational and labor market outcomes.[4] Altonji (1995) finds small effects for several high school courses in the US but specifically states the difficulties in estimating causal effects are too large to draw policy conclusions on the results. Several papers look at the impact of intensive math courses and found positive effects, at least for some groups of students (Rose and Betts (2004), Joensen and Nielsen (2009), Aughinbaugh (2012)). Papers that look at choices between academic

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

and vocational courses stress the importance of comparative advantages in different programs which causes heterogeneous effects (Kreisman and Stange (2017), Meer (2007)). I contribute to this literature by estimating the causal impact of multiple high school programs. I distinguish between four tracks that differ in their academic level and look at differences within tracks that prepare for higher education, based on the math-intensity of the curriculum and if classical languages are included. The benefit of the Flemish institutional context is that the study program does not have an impact on the higher education options students can choose from. This allows me to identify the effect of both academic and vocational study programs. I also estimate a model that explains how students end up in a certain program. This allows me to investigate policies that help to match students to their program, but also to calculate the added value of each program, i.e. the total impact of the availability of each program on grade retention, drop out and long run outcomes.

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

Finally, this paper contributes to the development and estimation of structural models of educational choices. Dynamic discrete choice models have often been used to evaluate the impact of counterfactual policies on choices of study programs.[5] Dynamics are important because students are expected to anticipate future policy changes by choosing other study programs. The same type of models have also been used to look at the impact of wage returns on program choices.[6] One of the downsides of these models is that they do not allow performance in the model to be a function of the choice of study effort. Nevertheless, theoretical (Costrell 1994) and reduced form (Garibaldi *et al.* 2012) evidence suggests that these dynamic incentives should have an impact. I contribute to this literature by adding study effort to the model, without requiring additional data or exclusion restrictions. By modeling

[5] See e.g. Eckstein and Wolpin (1999), Arcidiacono (2005), Joensen (2009) or Declercq and Verboven (2017).

[6] See e.g. Arcidiacono (2004) and Beffy *et al.* (2012).

study effort, I give a structural interpretation to the realization of performance outcomes. This differs from traditional structural models that assume end-of-year performance follows an exogenous law of motion.[7] While the data requirements and the number of estimated parameters are the same, there are important implications for counterfactual predictions because unobserved study effort is now allowed to react to a new policy. The estimation of dynamic discrete choice models also benefitted from methodological contributions of Rust (1987), Hotz and Miller (1993) and Arcidiacono and Miller (2011). I extend these methods to allow for effort choice such that estimation of the model remains feasible. I also compare my results to the standard model that assumes performance follows an exogenous law of motion.

I find that high school programs are important for long run outcomes. Without the academic track, the percentage of college graduates (39%) would decrease by 8.6 percentage points (%points). Elective courses matter mainly for the type of higher education degree. Study programs that include classical languages in the curriculum raise the number of graduates at the most prestigious higher education institutes, universities, by 1.9 %points and intensive math courses lead to more STEM majors: 2.5 %points. These are large numbers given that only 9.6% graduates from universities and 10.2% studies a STEM major. The availability of the vocational track reduces higher education outcomes, but it also decreases drop out rates by 10.7 %points and grade retention rates by 9.4 %points. It also helps in decreasing the average unemployment spell between age 25 and 35 of 2.08 years by one month.

These results show that different programs serve different outcomes. This is because different students are affected by the availability of each program. Tracking policies should therefore be designed carefully. In the evaluation of different tracking policies, I find that allowing underperforming students to switch tracks as an alternative for repeating a grade has important benefits in the long run. Without this, the percentage of students with grade retention would increase by 9.5 %points, or 1 out of 3 students, and college graduation would decrease by 1.8 %points. This suggests that underperforming students should be encouraged to switch to a program of lower academic level, rather than spending extra time in school to graduate from a more academic program. I also find that this policy can be further improved to avoid costly grade retention, without hurting student's long run outcomes. Prohibiting students to repeat a grade if they can avoid this by switching programs would decrease the number of students who were retained in secondary education by 9.6 %points. This does decrease graduation from higher education oriented tracks by 2.3 %points and enrollment in

[7]This type of models has been used mainly in the college major choice literature, see Altonji et al. (2016) for a recent overview.

higher education by 1.2 %points, but there are no negative effects in the long run. There is even a small decrease in the unemployment spell and a small, but insignificant, increase in obtaining a college degree.

Finally, I find that a model without effort choice underestimates the positive effects on student outcomes of both counterfactual simulations. This is because students increase their effort level to avoid being unqualified to continue in the program. This is especially the case for a policy that no longer allows students to avoid grade retention by switching programs. A model without effort would predict an increase in grade retention of 12.6 %points instead of 9.5 and a decrease in college graduation of 2.7 %points instead of 1.8. Also other outcomes show significant differences. Similarly, the estimated effects of the availability of each program are biased when we do not account for their impact on study effort.

The rest of the paper is structured as follows. Section 2 describes a dynamic model in which both study program and effort is chosen, how it relates to the literature and how it is identified. In Section 3 I explain how to estimate the model. Section 4 discusses the institutional context and introduces the data and section 5 applies the model to this context. I discuss the estimation results and the fit of the model in section 6, the added value of high school programs in section 7 and I evaluate tracking policies in section 8. Finally, I discuss the main limitations of the model in section 9 and conclude in section 10.

2 Dynamic model of program and effort choice

This section introduces a dynamic model of educational choices in which students simultaneously choose a study program and an effort level, taking into account the (psychic) cost of education and the future impact of their choice. I assume a finite horizon, limited by the year following the last year a student is allowed to study in secondary education: $t = T^{\max} + 1$. Long run outcomes are not modeled in a structural way. At the end of this section I explain how the effect of secondary education programs and counterfactuals on long run outcomes can still be estimated without a structural model.

Throughout the model, i refers to a student, t the time period in years and $j = 0, ..., J$ different study programs to choose from. These programs are mutually exclusive and are either programs in secondary education ($j \in se$), higher education ($j \in he$) or the outside option ($j = 0$). The outside option includes all other options like work, unemployment or following training programs outside of school. Not every option is always available as compulsory education laws prohibit $j \notin se$ until a certain age (18 in Belgium) and $j \in he$ requires a higher education degree. The choice set is given by Φ_{it}. The study program chosen by a student is denoted by a vector of dummy variables $d_{it} = (d_{it}^0, d_{it}^1, ..., d_{it}^J)'$ and $g_{it} = \{1, 2, ..., G\}$ is an end-of-year performance measure. This measure is assumed to be a discrete result students obtains at the end of each year with 1 the lowest possible outcome and G the highest. It will be modeled as function of unobserved study effort y. In the application this will take the form of the academic level the study program in the next grade is allowed to have during secondary education, and a high school degree at the end of secondary education. The information set of students is characterized by state variables x_{it} and ν_i and iid shocks at time t. The econometrician only observes x_{it}, the chosen programs d_{it} and the performance measures g_{it}, but not the unobserved (ability) type of a student ν_i or the iid shocks at time t. Each student belongs to one unobserved type $m = 1, ..., M$, indicated by a dummy variable equal to one inside the vector $\nu_i = (\nu_i^1, \nu_i^2, ..., \nu_i^M)'$. I will account for observed measures of language and math ability in x, but type is still expected to be important because these variables are expected to be noisy measures of the cognitive skills of a student and do not capture non-cognitive skills.

2.1 Defining effort

I define effort y as a sufficient statistic of the distribution of the discrete end-of-year performance measure $g_{it+1} = \{1, 2, ..., G\}$. Let performance in $t+1$ be the result of effort today y_{ijt} and an iid shock η_{ijt+1} such that:

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

where $\bar{\eta}_j^{\bar{g}}$ denotes the program-specific threshold to obtain at least outcome $\bar{g}$. Note the difference in timing between the chosen effort level and the shock. Students choose, and therefore know, y_{ijt} when they make a joint decision on a program and an effort level at time t. However, they do not know the realization of g_{it+1} because of the shock η_{ijt+1}. This shock captures uncertainty in grading standards or unexpected events during the year. The only information students have is the probability of obtaining outcome $\bar{g}$ in program j if they choose effort level y_{ijt}:

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

with $F(.)$ the cumulative distribution function of the distribution of the shock. Setting $\bar{\eta}_j^1 = -\infty$ and $\bar{\eta}_j^{G+1} = +\infty$ guarantees that all probabilities add up to 1. I assume η_{ijt+1} is logistically distributed such the probability of the each outcome can be written as follows:

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

Performance is then the result of an ordered logit model with index $\ln y_{ijt}$. Since (1) remains equivalent when adding or subtracting the same term on all sides, I can normalize one of the thresholds $\bar{\eta}_j^2 = 0$ such that all other thresholds should be interpreted with respect to the threshold of obtaining at least $g_{it+1} = 2$. Rewriting (2) for the lowest outcome ($g_{it+1} = 1$) shows that effort is defined as the odd of avoid the lowest outcome:

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

2.2 The value of each study program

Each year, students solve a dynamic problem.[8] They choose the study program j with the highest expected lifetime utility, in which effort y is chosen such that it maximizes the value of choosing that program. The value of each program can be represented by a Bellman equation:

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

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

The conditional value function is decomposed into $u_j(x_{it}, \nu_i, y_{ijt})$, the flow utility of schooling, and the discounted expected value of behaving optimally from $t + 1$ on, with $\beta \in (0, 1)$ the one-year discount factor.[9] 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 outcomes after high school like the value of going to college, wages and leisure in the future. It can be written as a weighted sum over the ex-ante value functions $\bar{V}_{t+1}(x_{i,t+1}, \nu_i)$, i.e. the value functions integrated over the iid shocks in state x_{it+1}. Since performance measure g

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

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

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

Traditional models of educational decision directly estimate the flow utility as a function of the state variables and estimate an exogenous law of motion on performance. I instead split up the flow utility in $C_j^0(x_{it}, \nu_i)$, the fixed cost students attach to schooling, and $c_j(x_{it}, \nu_i)y_{ijt}$, the variable effort cost:

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

This fixed cost includes cost components that are not associated with performance. This captures a (dis)taste to go to school, which is allowed to depend on student or family characteristics because of differences in preferences or social norms. Note that this cost can also be negative because students might enjoy going to school or parents can reward (or force) them to go to school. $c_j(x_{it}, \nu_i)$ is the marginal cost of effort and y_{ijt} the effort level the student chooses. The marginal cost captures that, conditional on future values, students dislike the effort that is required to perform better. 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. Note that this functional form implies a constant marginal cost of effort assumption. However, this does not imply that there is a constant cost to increase the probability to perform better. Effort is defined as the odd of avoiding the lowest performance outcome (see (3)). If the probability of the lowest outcome is already low, it is therefore much more costly to further decrease it, than if the probability is high.[10] The model is therefore consistent with the notion that increasing the probability of being successful in the program becomes more costly when this probability is already high.

2.3 The value of leaving secondary education

Each period t, students take two decisions: (1) which program j to study and (2) which effort level y to exert. The model can be solved by backward induction, provided we know the expected value of the lifetime utility after leaving secondary education $t = T_i^{SE} + 1$, and we know the final year $t = T^{\max}$ in which students are allowed to go to secondary education. I therefore impose $T^{\max}$ and estimate the lifetime utility of leaving secondary education with and without a degree.[11] To derive the value of a high school degree, I look

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

[11] In Flanders, secondary education takes 6 years to complete. I impose $T^{\max} = 9$, implying students can have at most three years of grade retention and are no longer able to be in secondary education in the 10th year.

at the choice they make in the year after leaving secondary education. Students with a high school degree can choose among higher education options $j \in he$ or the outside option $j = 0$. The outside option captures (un)employment or training programs outside of the regular system of tertiary education Students without a high school degree can only choose $j = 0$.

Let the value of each option after leaving secondary education ($j \in \{he, 0\}$ and $t = T_i^{SE} + 1$) be given by

$$v_{ijt}(x_{it}, \nu_i) + \varepsilon_{ijt} = \text{Degree}'_{it}\mu^{\text{degree}} + \Psi_j^{HEE}(x_{it}, \nu_i) + \varepsilon_{ijt} \tag{6}$$

with $\text{Degree}'_{it} \subset x_{it}$ a vector of dummy variables for each possible degree a student can obtain in high school, μ^{degree} a vector of parameters to estimate and $\Psi_j^{HEE}(.)$ a reduced form function of the state variables that predict the higher education enrollment (HEE) decision. μ^{degree} captures the fact that students stay in school to obtain a degree, also if they do not go to higher education. $\Psi_j^{HEE}(.)$ describes how state variables affect the lifetime utility of each option. Since only differences in utility are identified, I normalize $\Psi_0^{HEE} = 0$. Note that the common parameter μ^{degree} can still be identified by choice behavior during secondary education. This is because of the dynamics of the model. Students who are more likely to graduate because of higher expected performance, or because they have almost completed all grades of high school, will be less likely to drop out. This is similar to Eckstein and Wolpin (1999) who use data in secondary education to identify the value of graduating from high school.

2.4 Solving the model

It is now possible to solve the model by backwards induction. In the last period, secondary education is not possible and the model can be solved like a static one. Because I assume $h(\varepsilon_{it})$ is extreme value type 1, I can write the expected value of lifetime utility in the period where secondary education is no longer allowed as follows:

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

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

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

The optimality condition is a standard equalization of marginal costs and marginal benefits and allows us to write marginal costs as a function of future values and the way effort changes performance. Note that we implicitly assumed effort is only affecting end-of-year performance but not the value of behaving optimally in the future, conditional on performance. Therefore we only need to take into account the derivative of the performance measure and not how future values would react directly to a change in effort. In appendix section A.1 I show that the marginal benefits are always positive and decreasing in effort y. They follow an S-shaped curve, bounded by 0 and a weighted sum of the gains of obtaining a better performance measure.

For a binary performance measure ($G = 2$), optimal effort has a simple analytic solution with an intuitive interpretation:

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

Optimal effort in period $t-1$ increases in the discounted benefits of obtaining performance level 2 instead of 1 in period t, and decreases in marginal costs c in period $t-1$. This shows a clear dynamic trade-off. Extra effort at time $t-1$ is costly but generates benefits in t. The way effort was defined in section 2.1 restricts the domain of y to be in the open set $(0, +\infty)$. An interior solution is therefore required. We see that this assumption puts an upper bound on marginal costs. If marginal costs of effort 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 there is more than one performance measure (see appendix section A.1).[12]

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

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

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

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

This procedure can be repeated until the start of secondary education to solve the entire model.

2.5 Long run outcomes

To evaluate the impact of high school programs on long run outcomes, a structural model in secondary education is needed as it allows for policy counterfactuals that will not change

[12] A sufficient behavioral condition that implies this model assumption is to assume that a student always believes there is a non-zero probability of avoiding the worst performance outcome.

the primitives of the model, like fixed costs, marginal cost of effort or the value of a degree, but it will change student behavior. Without a structural model, we would not be able to assess the effects of changes in policy. For outcomes after secondary education, we do not need to know the primitives of the model but only the way these outcomes are influenced by secondary education outcomes, after controlling for observed and unobserved student characteristics. I therefore model a reduced form function only. Let $w \in W$ be an outcome variable that can be described by a function of the state variables (x_{it_w}, ν_i) and an iid shock ω_i^w:

$$\tilde{\Psi}^w(x_{it}, \nu_i, \omega_i^w) \text{ if } t = t_w \tag{8}$$

with t_w the time period when the long run outcome w is realized and x_{it_w} containing characteristics of students, but also the high school program from which the student graduated and the years of study delay he accumulated in secondary education. Correlation between long run outcomes and other parts of the model is captured by observables x and unobserved type ν and is crucial to avoid ability bias in the estimates. We also need to control for long run outcomes in x_{it_w} that are realized at $t < t_w$ to allow for ω_i^w to be modeled credibly as an iid shock.[13] Other assumptions on $\tilde{\Psi}^w(.)$ and ω_i^w will be imposed when applying the model to the institutional context in section 5.

After estimation, we can use these functions to look at the impact of counterfactual policies in secondary education. Let $x_{it_w}(Policy = 0)$ be the realized state vector of i at time t_w in the status quo scenario, and $x_{it_w}(Policy = p')$ the state vector in the counterfactual scenario. The expected effect on long run outcome w of policy p' is then given by:

$$E_{x,\nu,\omega}\left[\tilde{\Psi}^w(x_{it_w}(Policy = p'), \nu_i, \omega_i^w) - \tilde{\Psi}^w(x_{it_w}(Policy = 0), \nu_i, \omega_i^w)\right]$$

with $E_{x,\nu,\omega}$ an expectations operator over the empirical distribution of the observables x and the estimated distribution of the unobserved types ν and shocks ω.

2.6 Relation to the literature

Most study choice models use a special case of the value function in (4) to estimate the model. A dynamic model ($\beta \in (0, 1)$) is often needed to capture the fact that students make different study program choices in anticipation of policy changes that will affect their utility in the future. Similarly, (wage) returns to education are expected to influence these program choices. Although these models are dynamic, they all set marginal costs $c_j = 0$ and either do not model y (Keane & Wolpin 1997) or assume it is exogenous, conditional on observed

[13] Take for example a shock that causes students to graduate from higher education. This is also expected to have an impact on labor market outcomes through its effects on higher education.

decision variables. With $c_j = 0$, y only enters the conditional value function through its impact on future benefits and can therefore not be treated as a choice variable as students would always exert maximum effort. In practice, researchers estimate law of motions that depend on study program choices and on past results. Examples are grade equations in Eckstein and Wolpin (1999) and Arcidiacono (2004), course credit accumulation in Joensen (2009) and Declercq and Verboven (2017), college admission probabilities in Arcidiacono (2005) or length of study in Beffy *et al.* (2012). In this type of models, counterfactual simulations allow students to choose a different program j, but not a different effort level y, conditional on the program choice and the realization of the state variable. Law of motions are estimated reduced form or recovered nonparametrically from the data, and kept constant in these simulations. This is a strong assumption as most policies that are expected to change program choices because of dynamic considerations, are also expected to change effort levels. An increase in the return to better grades, more credits, college or length of study is not only going to make some programs more popular, it will also induce students to exert more effort to make sure they graduate from them.

There is also a related literature on structural models of job search. Some models endogenize the probability to find a job, or search intensity, by equating marginal costs and marginal benefits (Paserman (2008), van den Berg *et al.* (2015) and Cockx *et al.* (2017)). This paper applies a similar identification strategy for marginal costs, but within a rich model of educational decisions. The richness of the model comes at the cost of computational complexity. In section 3 I describe how this burden can be reduced significantly.

2.7 Identification

I first discuss identification if the type of a student, ν_i, is observed by the econometrician and then discuss how types can be identified if they remain unobserved. As is common for dynamic discrete choice models, identification of flow utility parameters depends on the distributional assumptions on iid shocks ε_{ijt}, only differences in utility are identified and require the normalization of the flow utility of one option ($j = 0$) (Magnac & Thesmar 2002). The discount factor β is set before estimation.[14] The added complexity in this paper, is that I split up the flow utility of a study program in two components that depend on the state variables: a fixed cost $C_j^0(x_{it}, \nu_i)$ and a variable effort cost as a function of marginal costs $c_j(x_{it}, \nu_i)$. Nevertheless, I do not estimate more parameters than a model that does not make this distinction. This is because other models estimate the law of motion of performance measures directly, instead of assuming it is generated by the structural model. Instead of using

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

data on program choices and measures of performance to estimate flow utility and a law of motion, I use it to estimate a component that is independent of performance: fixed costs, and a component that rises in the probability to perform better: marginal costs. Optimal behavior implies that marginal costs can be identified from marginal benefits that arise naturally in a dynamic model (see equation (7)). This avoids the need for exclusion restrictions. The following example shows how data on program choices and data on performance identify different parameters. Take two students with the same characteristics that have the same probability to go to each college option after high school, conditional on their characteristics and high school background. If we observe them in their last year of high school in the same program but with different distributions of performance, then the one with worse performance must have a higher marginal cost of effort, regardless of their fixed costs. If two students are equally likely to go to college, conditional on their high school background and they have the same performance distribution, they must have the same marginal costs (see equation (7)). Fixed costs can then be identified by looking at differences in program choice or drop out the year before.

I now turn to the identification of the unobserved types. Note that all shocks in the model are assumed to be iid. This holds for the flow utility shocks ε_{ijt}, performance shocks η_{ijt} and shocks on long run outcomes ω_i^w. Any correlation we see in the data that cannot be captured by observable characteristics, will therefore help in identifying the unobserved type. For example, when we observe two students with identical observable characteristics but one consistently outperforms the other during high school, it reveals something about the student's unobserved ability that will help him in high school, but might also help him in higher education. A second source of identification are exclusion restrictions in the model. In particular, I will assume that travel time to high school options influences selection into programs, but has no direct effect on outcomes after secondary education.[15] Similarly, distance to higher education options does not influence labor market outcomes directly. This instrumental variables strategy helps in separately identifying the unobserved types from the effect of a high school program. If students living nearby specific programs obtain better outcomes, the model must attribute this effect to the program and not to unobserved ability because unobserved ability is assumed to be uncorrelated with distance and time, while distance and time do influence program choices.

[15] See De Groote and Declercq (2017) for a discussion on the validity of distance to school as an instrument for school choice in this context. The paper assumes that distance to school is uncorrelated with success in high school. Here I make a weaker assumption by saying it is uncorrelated with success after high school. I also make an additional assumption, saying that there is no direct effect of distance on choices after high school, after controlling for distance to higher education.

3 Estimation

This model can be matched to the data, using maximum likelihood. This estimates all parameters of the model, after imposing functional form assumptions on fixed costs $C_j^0(.)$, marginal costs $c_j(.)$ and the reduced form functions. At the same time, a distribution of types ν can be estimated. However, solving the model and estimating the parameters at the same time is computationally costly. Even in models without a choice of effort, many dynamic models on educational decisions have relied on methods that avoid solving the entire model when estimating parameters (see e.g. Arcidiacono *et al.* (2016), Declercq and Verboven (2017) and Joensen (2009)). With endogenous effort, the computational issues increase further because state transitions that follow from performance can no longer be estimated in a first stage as performance is an endogenous function of the choice of effort. Nevertheless, I can show that solutions to avoid fully solving the model, pioneered by Hotz and Miller (1993) and extended to allow for persistent unobserved heterogeneity ν_i in Arcidiacono and Miller (2011), can still be applied.

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

3.1 Estimation when student type is observed

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

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

For $j \in \{0, he\}$ we can substitute in (6), which is a function of observed variables and parameters to estimate. For $j \in se$ this is not possible as (4) does not only depend on a function of parameters to estimate, but also on optimal behavior in the future and the optimal level of effort. We would therefore have to solve the model as explained in section 2.4 to be able to estimate all parameters. To reduce the computational burden, I avoid solving the model during estimation. I first show how the first-order condition (7) allows me to substitute out the marginal cost function, when estimating other parameters of the model and how the optimal level of effort can be derived from the data. I can then apply the CCP method of Hotz and Miller (1993) to avoid solving the dynamic model during estimation. Finally I show how additivity of the likelihood function allows for each equation in the model to be estimated separately.

Optimal effort

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

$$y^*_{jt}(x_{it}, \nu_i) = \frac{1 - P_j(g_{it+1} = 1|t, x_{it}, \nu_i)}{P_j(g_{it+1} = 1|t, x_{it}, \nu_i)}. \tag{10}$$

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

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

$$\begin{aligned} & v_{ijt}(x_{it}, \nu_i, y^*_{ijt}) + \varepsilon_{ijt} \\ = \; & -C^0_j(x_{it}, \nu_i) \\ & + \beta \sum_{\bar{g}} \left[\bar{V}_{t+1}(x_{it+1}(\bar{g}), \nu_i) \left(P_j(g_{it+1} = \bar{g}|y^*_{jt}(x_{it}, \nu_i)) - \frac{\partial P_j(g_{it+1} = \bar{g}|y_{ijt})}{\partial y_{ijt}}|_{y_{ijt}=y^*_{ijt}} y^*_{jt}(x_{it}, \nu_i) \right) \right] + \varepsilon_{ijt} \end{aligned} \tag{11}$$

with $y^*_{jt}(x_{it}, \nu_i)$ given by (10). $\frac{\partial P_j(g_{it}=\bar{g}|y_{ijt})}{\partial y_{ijt}}$ can be derived from the distributional assumptions on the performance measure. The conditional value function is now written with the same unknowns as in standard dynamic models with exogenous state transitions, following Rust (1987). The only difference is the transition matrix. This matrix characterizes how current states impact utility in the future. In a model without effort, this depends only on how states transition in the data. Since effort now offers a way to increase these outcomes, it now also depends on y and how it affects state transitions.

CCP method

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

$$\bar{V}_{t+1}(x_{it+1}, \nu_i) = \gamma + v_{id^*t+1}(x_{it+1}, \nu_i) - \ln \Pr(d^*_{it+1}|x_{it+1}, \nu_i) \tag{12}$$

with d^*_{it+1} the vector of dummy variables for each option in which the indicator of one arbitrary option is set to 1 and $v_{id^*t+1}(.)$ the conditional value function of this option. This is particularly useful when there is an option that terminates the (structural) model. If it is possible to leave secondary education in $t+1$, we can choose $j=0$ as the arbitrary choice and substitute its value function (6) in (12):

$$\overline{V}_{t+1}(x_{it+1},\nu_i) = \gamma + \text{Degree}'_{it}\mu^{\text{degree}} + \Psi_0^{HEE}(x_{it+1},\nu_i) - \ln \Pr(d^0_{it+1}=1|x_{it+1},\nu_i). \quad (13)$$

As explained in section 2.3, $\Psi_0^{HEE}(.)=0$ by normalization. We can now substitute (13) in (11), such that for all $j \in se$:

$$\begin{aligned} & v_{ijt}(x_{it},\nu_i,y^*_{ijt}) + \varepsilon_{ijt} \quad (14) \\ = \quad & -C^0_j(x_{it},\nu_i) + \beta\gamma \\ & +\beta \sum_{\bar{g}} \left[\begin{array}{c} \left(\text{Degree}'_{it}(\bar{g})\mu^{\text{degree}} - \ln \Pr(d^0_{it+1}=1|x_{it+1}(\bar{g}),\nu_i)\right) \\ \left(P_j(g_{it+1}=\bar{g}|y^*_{jt}(x_{it},\nu_i)) - \frac{\partial P_j(g_{it+1}=\bar{g}|y_{ijt})}{\partial y_{ijt}}|_{y_{ijt}=y^*_{ijt}} y^*_{jt}(x_{it},\nu_i)\right) \end{array} \right] + \varepsilon_{ijt}. \end{aligned}$$

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 2.3). As in Hotz and Miller (1993), a nonparametric estimate of the Conditional Choice Probability (CCP) $\Pr(d^0_{it+1}=1|x_{it+1},\nu_i)$ can be recovered from the data, before estimating the model.[16] Because of compulsory education laws, the outside option is not always in the choice set. In the appendix, I show how the concept of finite dependence, introduced in Arcidiacono and Miller (2011), can be used to overcome this problem without fully solving or simulating the model during estimation.[17]

Likelihood function

We can now use (14) to estimate the fixed cost parameters of $C^0_j(.)$, a functional form for long run outcomes $\tilde{\Psi}^w(.)$ (including (6) for $w=HEE$) and the common component of a value of a high school degree μ^{degree}. This can be done by using the probabilities according to the model and data on program choices and long run outcomes. Let the fixed cost parameters in $C^0_j(.)$ and the common value of a degree μ^{degree} be given by μ and the reduced form parameters by ϕ. Assuming iid observations, the loglikelihood of the data is[18]:

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

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

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

[18] Note that data on performance measures does not enter the likelihood function because its distribution is already recovered directly from the data. In the application of the model I will however approximate the optimal level of effort by a parameteric function such that the performance measure also enters the likelihood function.

with $L_{it}^{program}(\mu, \phi)$ given by logit choice probabilities (9) and $L_i^w(\phi)$ given by the assumed processes on long run outcomes (8). This loglikelihood could be maximized directly to obtain the estimates of (μ, ϕ). However, because of additive separability, consistent estimation could also be performed in sequential steps by first estimating the process of each long run outcome and then estimate the structural model. After estimation, the marginal cost function $c_j(.)$ can be recovered from the first-order condition (7) without requiring any structure on its functional form.

3.2 Estimation when student type is unobserved

To allow for persistent unobserved heterogeneity, I follow Arcidiacono and Miller (2011) and estimate a finite mixtures 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.[19] The loglikelihood function then becomes:

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

There are three main changes to the likelihood function. First, $\pi_{m|age\_start}$ is added as an parameter to estimate, specific for each starting age, and likelihood contributions are conditioned on the type. Second, the likelihood contribution of the performance outcome in secondary education $L_{it+1}^{\text{performance},m}$ is added to the likelihood function. This is needed to recover the optimal levels of effort in the data as they depend on the unobserved type. Therefore they can no longer be recovered from the data before estimating other parameters. Third, the function is no longer additively separable such that sequential estimation is not possible anymore.

Additive separability can be restored using the estimator of Arcidiacono and Miller (2011). The estimation procedure is an adaptation of the EM algorithm. The algorithm

[19] 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.

starts from a random probability of each observation to belong to each type. The entire model can then be estimated as explained in section 3.1, but weighs each observation-type combination by the probability that the observation belongs to the type. Afterwards, the joint likelihood of the data conditional on each type, is used to update the individual type probabilities with 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 optimal effort levels in the data $y_{jt}^*(x_{it}, \nu_i)$ and the thresholds $\bar{\eta}_j$ are identified in a first stage.[20] In a second stage, the fixed cost parameters and the common component of the value of a degree μ can be recovered using the structural model. Finally, the first-order condition (7) is used to recover the marginal costs. Standard errors can be obtained by using a bootstrap procedure.[21]

4 Institutional background and data

Before applying the model to the data, I first describe the institutional context that is relevant for the model and the counterfactual analysis. I also introduce the dataset and discuss some descriptive evidence. I make use of the LOSO dataset in which I follow a sample of 4927 students in Flanders (Belgium) that started secondary education in 1990.[22] I follow them through their entire career in secondary education but also up to 18 years afterwards, allowing me to observe the long run outcomes.

4.1 Study programs

Belgium is a federal country, divided in communities and regions. The communities have jurisdiction over all educational affairs. There is a Dutch-speaking Flemish Community, a

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

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

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

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

French Community and a German Community. I discuss the schooling system of the Flemish Community (which represents about 60% of the Belgian population).[23]

After finishing six grades in elementary school, students enroll in secondary education (high school) in the 7th grade, usually in the calendar year they become 12 years old. Students can choose between all schools in Flanders since school choice is not geographically restricted and free school choice is law-enforced.[24] In practice, most students choose one of the closest alternatives. After high school, students can enroll in higher education.

Students in full time education choose between different high school programs, grouped into tracks that differ in their academic level.[25] 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.[26] Therefore I follow Baert *et al.* (2015) 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, that prepares them for specific occupations that do not require a higher education degree. Within each track, students can choose several programs which consist of bundles of elective courses. Since there are many programs to choose from, I aggregate them up to 8 study programs. In particular, I split up the academic track in four programs: classical languages, intensive math, intensive math + classical languages and other. The middle-theoretical track is split between intensive math and other.[27] This aggregation still allows for a sufficient number of students in each group and corresponds to important differences in enrollment and success rates in higher education (Declercq & Verboven 2015).[28]

A student graduates from high school after a successful year in the 12th grade in the academic or one of middle tracks, or the 13th grade in the vocational track. Compulsory education laws require a student to pursue education until June 30th of the year he reaches

[23] Almost all schools in Flanders either belong to the 'official education' or to 'free education'. Although only the official education institutions are government-owned, also the 'free education' institutions have to conform to educational standard set by the Flemish government. Moreover, they are all subsidized such that there are no tuition fees and the financial cost of education is limited to the cost of personal school material and school trips.

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

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

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

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

[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 will not distinguish between different schools as they are all regulated in the same way and the restrictions implied by certificates also hold for other schools.

the age of 18. From the age of 15, he can also decide to leave full time education and start a part-time program in which he can combine working and schooling.[29]

Despite of the fact that each track prepares for different options after secondary education, enrollment in almost any higher education option is free of selection by track or by the higher education institutes themselves. Students from any track can enroll in almost any program of higher education (Declercq & Verboven 2017). Therefore, selection into higher education only takes the form of self-selection.

4.2 Tracking policy

At the start of secondary education, all programs are available. The choice set in the future depends on the current program and performance during the year. Upward mobility, i.e. moving from a track of lower academic level to a more rigorous one, is practically impossible, except for switches between middle tracks and the academic track in the first two grades. Similarly, students can never enroll in programs with classical languages anymore if they did not choose it from the start. Math-intensive programs are available from grade 9 on. From then on, switching from a program without extra math to one with intensive math in the academic track is not possible. Similarly, students in the middle track that did not have extra hours of math cannot choose this anymore. Finally, there can be no more switching between full-time programs from grade 11 on.[30]

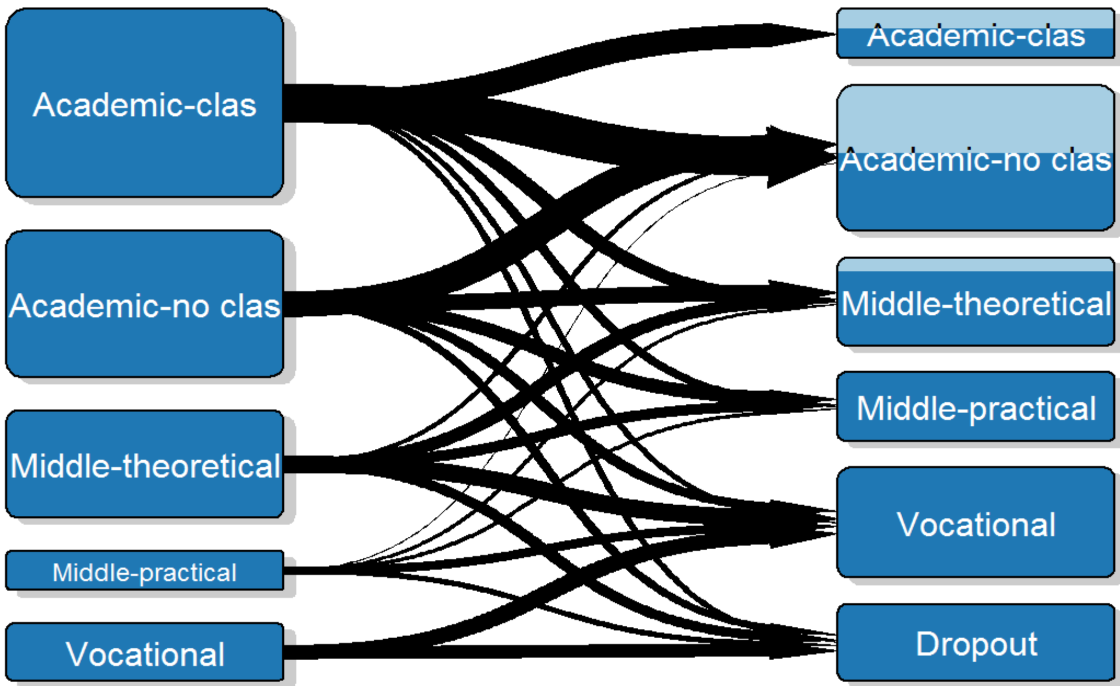
Performance also matters for the choice set. Each year, students obtain a certificate. An A-certificate means the student succeeded on all courses. He can then move on to the next grade and continue in the program. If he did not, teachers decide on the certificate he gets. This can still be an A-certificate, e.g. if the student only failed on a small number of courses, but it can also be a B- or a C-certificate. A C-certificate means that the student failed on too many important courses and has to 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 is allowed to proceed to the next grade, but he will be excluded from some programs, specified in the certificate. Alternatively, a student with a B-certificate can decide to repeat the grade without being excluded from a program. In most cases, a B-certificate excludes the track a student is currently in and therefore encourages them to downgrade to another track. However, a B-certificate can also exclude only certain elective courses within a track (see appendix Table A25).

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

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

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

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



Note: Left: program chosen in grade 7, right: last choice before leaving secondary education. Clas= classical languages included. Light blue area = proportion of students in program with extra math. See appendix Table A26 for data on these transitions.

These transitions are not always a smooth or voluntary process. Each year students obtain a certificate that can restrict their choice set. C-certificates prohibit students from going to the next grade. B-certificates can cause grade retention if students do not want to switch to a different program. Most of the time students obtain an A-certificate. However, 8% of the certificates are B-certificates and 7% are C-certificates. One out of four students with a B-certificate also decides to repeat the grade instead of downgrading, i.e. switching to a track of lower academic level or dropping an elective course. 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. This causes a lot of grade retention. Table 1 shows that 33% of students leave high school with at least one year of study delay. There is a large difference between

Table 1: Performance during secondary education

	At least 1 B-certificate (% of students)	**At least 1 C-certificate** (% of students)	**At least 1 year of study delay** (% of students)	**Obtains high school degree** (% of students)
All	38.1	30.7	33.2	78.3
Conditional on first choice of study program				
Academic	34.3	23.5	29.0	91.3
Middle-Theoretical	60.0	37.6	43.5	72.7
Middle-Practical	66.4	44.0	41.2	54.6
Vocational	0.0	50.8	33.3	28.2
Conditional on last choice in secondary education				
Academic	10.6	11.4	15.3	100.0
Middle-Theoretical	41.9	24.8	33.2	100.0
Middle-Practical	55.8	36.3	51.6	100.0
Vocational	62.8	32.8	36.4	63.0
Drop out	52.6	78.7	58.3	0.0

Note: Different measures of performance, conditional on the study program chosen at the start of secondary education and on the final choice. A-certificate: proceed to next grade, C-certificate: repeat grade, B-certificate: repeat or downgrade. Degree obtained after successfully completing the 13th grade in the vocational track or the 12th grade in other tracks.

the performance measures of students who started in the academic track if we compare them to students who also graduated from the academic track. 15% of students who graduated from the academic track obtained at least one year of study delay, but students who started in it are twice as likely to be retained. This is reflected in the certificates they get. They are three times as likely to obtain a B-certificate and two times as likely to obtain a C-certificate. This shows that track switching is not just a choice of students but the certificate policy has an important impact on steering choices.

4.3 International comparison

The Flemish educational system resembles mostly the early tracking systems in Europe, like Germany, Austria or the Netherlands, because it sorts students on ability at an early age. However, it also shares characteristics of more comprehensive educational systems by allowing free track choice at the start of secondary education, having many schools with multiple tracks and allowing access to all higher education options from any track. This setting provides an important application of the model and is particularly useful to study the impact of high school curriculum and tracking policies.

First, free access to all higher education options allows me to identify the impact of all study programs, including vocational programs, on higher education outcomes. Second, there is a uniform tracking policy across the region to encourage students to choose the program that matches best with their preferences and ability. The B-certificate is of particular interest. While many countries offer ways to avoid grade retention by changing program, there is not always a mechanism across schools that we can use to estimate a behavioral model.[31] The implications of this type of trade-off between grade retention and studying a rigorous program is not only interesting for countries that separate students in different tracks. Also countries like the US that differentiate more at the course-level can require students to retake failed classes. Moreover, many universities explicitly ask for a high GPA and a rigorous academic curriculum in their admission criteria. Students, especially those of lower ability, then face a similar trade-off between studying advanced courses at the risk of retakes and a lower GPA, or choosing a curriculum with less advanced courses.

Although similar issues arise in other educational systems, they are particularly important in the current context. Belgium spends 2.8% of its GDP on secondary education, the highest number among OECD countries. It is therefore crucial to study the effectiveness of the system in helping students to achieve their future goals in a cost-efficient way. Since 96% of the cost is paid for by government spending, it is also crucial to see if students have the right incentives within the system to optimize total welfare (OECD 2017). Finally, Belgium has a very high rate of grade retention in secondary education which comes at a large cost. The total cost of a year of study delay in Belgium amounts to at least 48 918 USD (corrected for PPP)/student or 11% of total expenditures on compulsory education, the highest percentage in the OECD (OECD 2013).

[31] The trade-off between grade retention and studying a program of higher academic level is also common in countries like Germany, Portugal, Lithuania, Luxembourg, the Netherlands, Austria, Liechtenstein and Slovakia (Eurydice European Unit 2011).

4.4 Study program and student background

Table 2 summarizes student characteristics by the program in which they graduate. The dataset contains measures on ability at the start of secondary education in the form of item response theory (IRT) scores based on standardized tests. These scores measure language (Dutch) and math ability and are standardized to be mean 0 and standard deviation 1. To capture differences in preferences for each program, I also include gender and socioeconomic status (SES) in the analysis. SES is measured by a dummy equal to one if at least one of the parents has completed higher education.

Not surprisingly, students with high SES end up much more in the academic track. Also initial ability matters. The average language ability of a student in the academic track is 0.69 standard deviations higher than the overall average. For students in the vocational track this is 0.73 below the overall average. Math ability shows a similar trend.[32] Moreover, we also see large differences within tracks with both classical languages and math programs attracting stronger students. Finally, gender is important too. Male students are less likely to be in tracks that prepare better for higher education, except for math-intensive programs.

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

[32] Note that this is very similar to other tracking countries. In Germany (grade 9), average language and math ability is 0.8 standard deviations above the average in the high track and 0.9 standard devations below the average in the vocational track (Dustmann *et al.* 2017).

Table 2: High school program and student background

Study program	Students		Male	Language ability	Math ability	High SES
All	*4927*	*(100.0%)*	*0.49*	*0.00*	*0.00*	*0.27*
Academic	*1787*	*(36.3%)*	*0.39*	*0.69*	*0.62*	*0.47*
clas+math	184	(3.7%)	0.46	1.09	1.02	0.62
clas	272	(5.5%)	0.35	0.91	0.65	0.57
math	621	(12.6%)	0.47	0.77	0.76	0.49
other	710	(14.4%)	0.32	0.44	0.38	0.38
Middle-Theoretical	*807*	*(16.4%)*	*0.54*	*0.13*	*0.21*	*0.22*
math	126	(2.6%)	0.71	0.33	0.48	0.32
other	681	(13.8%)	0.51	0.09	0.16	0.21
Middle-Practical	*626*	*(12.7%)*	*0.51*	*-0.04*	*0.01*	*0.23*
Vocational	*1009*	*(20.5%)*	*0.52*	*-0.73*	*-0.70*	*0.10*
13th grade	636	(12.9%)	0.50	-0.64	-0.64	0.12
12th grade	373	(7.6%)	0.54	-0.87	-0.80	0.07
Dropout	*698*	*(14.2%)*	*0.66*	*-0.87*	*-0.77*	*0.08*
Part-time	389	(7.9%)	0.70	-0.90	-0.80	0.07
Full time	309	(6.3%)	0.62	-0.82	-0.74	0.08

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

4.5 Higher education and unemployment

Table 3 summarizes differences in the main higher education outcomes, conditional on graduating from different study programs in secondary education. When students leave high

Table 3: High school program and long run outcomes: summary statistics

	Higher education			**Unemployment**
Study program	Enrollment (% of students)	First year success (% of enrolled)	Degree age 25 (% of students)	Spell age 25-35 (mean in years)
All	*55.2*	*45.6*	*39.0*	*2.08*
Academic	*94.2*	*52.8*	*76.7*	*1.43*
clas+math	97.8	63.3	87.0	1.02
clas	97.4	55.8	82.4	1.39
math	94.8	55.0	80.8	1.34
other	91.5	46.9	68.2	1.62
Middle-Theoretical	*79.3*	*38.7*	*48.1*	*1.65*
math	93.7	44.1	67.5	1.39
other	76.7	37.7	44.5	1.70
Middle-Practical	*50.2*	*30.6*	*23.8*	*1.69*
Vocational	*8.2*	*9.1*	*1.5*	*2.61*
13th grade	13.1	14.5	2.4	2.33
12th grade				3.10
Dropout				*3.86*
Part-time				3.95
Full time				3.74

Note: Clas= classical languages included. Math= intensive math. Students in vocational track only obtain full high school degree after an additional 13th grade. Drop out split between students directly opting for full time dropout or first choosing part-time option.

school, 55% decides to go directly to higher education.[33] Although not all tracks prepare

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

for higher education, they do not restrict any options. Therefore, the first year of higher education in practice serves as an admission test as overall success rates are very low and a lot of students do not finish the program they started (Declercq & Verboven 2017). Indeed, we see a first year success rate of 46% among enrolled students and only 39% of all students eventually obtains a higher education degree (by the age of 25). I also measure the years of unemployment between age 25 and 35 to look at the effects on labor market outcomes.[34]

There are large differences in higher education outcomes. 94% of all students graduating from the academic track start higher education, and the large majority eventually obtains a degree. For the other tracks this is less common. Nevertheless, the track distinction is not always clear from the results as students with extra math in the middle-theoretical track obtain similar outcomes as students in the academic track without extra math or classical languages. Despite the fact that the middle-practical track prepares primarily for the job market, we do see half of the students enrolling in higher education and 24% obtaining a degree.

Differences within the academic track become more clear when we look at different higher education choices. To capture heterogeneity in returns to high school study programs, I distinguish between different levels of higher education and different majors (see Table 4). Similar to Declercq and Verboven (2015), I distinguish between three types of higher education, increasing in their academic level (or prestige): professional college, academic college and university. In addition, I also distinguish between two majors in each level: STEM and other.[35] Most students graduate from professional colleges and only 1 out of 4 students with a higher education degree graduated from a STEM major. Only students who graduated from the academic track with classical languages in their curriculum are more likely to graduate from universities than from professional colleges. Extra math in the program is associated with graduating from STEM majors and academic colleges.

[34] Unemployment captures all years in which the individuals are not studying or working, regardless of the reason why they do not work.

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

Table 4: High school program and level and major college degree: summary statistics

	Academic level higher education			**Major**
	University	Academic college	Professional college	STEM
Study program	(% of students)	(% of students)	(% of students)	(% of students)
All	*9.6*	*4.8*	*24.8*	*10.2*
Academic	*25.8*	*10.2*	*40.6*	*18.4*
clas+math	56.5	14.1	16.3	33.2
clas	41.5	7.7	33.1	11.4
math	29.5	15.6	35.7	32.4
other	8.6	5.5	54.1	4.9
Middle-Theoretical	*1.0*	*4.7*	*42.4*	*17.7*
math	5.6	17.5	44.4	42.1
other	0.1	2.3	42.0	13.2
Middle-Practical	*0.5*	*1.9*	*21.4*	*4.6*
Vocational (13th grade)	*0.0*	*0.2*	*2.2*	*0.3*

Note: Three types of higher education options in decreasing order of academic level: university, academic college, professional college. Graduation rates adds up to the total graduation rate of 39.0%. Each level has different programs, aggregated to STEM and other majors, only STEM major is reported. Clas= classical languages included. Math= intensive math.

5 Application of the model

I now apply the model to the Flemish educational context. I specify the choice set of study programs, functional form assumptions of fixed costs of schooling, the performance measure and the long run outcomes of interest.

As explained in section 2, i refers to a student, t the time period in years and $j = 0, ..., J$ different study programs to choose from. The choice set is given by Φ_{it} and the study program chosen by a student is denoted by a vector of dummy variables d_{it}. g_{it} is an end-of-year performance measure, which is a function of the chosen effort level y. The information set of students is characterized by a time-varying observed state variable x_{it}, iid shocks at time t and a time-invariant type ν_i. The econometrician only observes x_{it}, the chosen programs d_{it} and the performance measure g_{it}.

5.1 Choice set

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

The program choices are restricted. First, students can never upgrade tracks according to the following hierarchy: $acad > midt > midp > voc > part$, with the exception of the first two grades in which mobility between $acad$, $midt$ and $midp$ is allowed. Second, they can stop with their specialization in extra math or classical languages, but the reverse is not possible.[36] Finally, from grade 11 on, students who want to stay in full time education must stay in the same program.

Students progress in secondary education by obtaining a certificate at the end of the year. As explained in the institutional context, the flexibility of a B-certificate can have different implications on the choice set. I therefore use the certificate data to create a variable that captures the permission for a student to enter or continue in each track in the next grade as a measure of performance g_{it+1}^{track}. To allow for B-certificates to only exclude elective courses, I extend the model to allow for additional performance measures that contain permissions to study classical languages g_{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 $j = he$ enrollment in a higher education (if they obtained a high school degree) and $j = 0$ the outside option. I assume this is a terminal choice, i.e. they never return to secondary education.

5.2 Costs and performance

Section 2 describes the model without specifying the variables that are used in the analyses and how they impact the schooling costs. The estimation section 3 explains the need for specifying a functional form for fixed costs, but marginal costs can be derived from the

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

marginal benefits at the optimal level of effort in the data. In this section I impose functional form assumptions on fixed cost and explain the performance measure that is used to derive the optimal level of effort from the data. I also extend the model to allow for more than one measure of performance such that B-certificates can also exclude elective courses.[37]

Fixed cost of secondary education

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

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

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 by bike to the closest school that offers the study option in the current grade and grade_{ijt} is the grade a student is in.[38] This first line allows for each program to differ in fixed costs and a linear interaction with the grade. Observed and unobserved characteristics of students influence the cost of each study program and a common cost of schooling is also allowed to change over grades by student characteristics. The second line allows costs to react to two different measures of grade retention, contained in the 2x1 vector: retention_{ijt}. This vector contains a flow variable: a dummy equal to one if the student is currently in the same grade as the year before and a stock variable that captures the years of study delay accumulated in previous years. The disutility of grade retention can differ over grades and over level_SE_{ijt} and the ranking of the academic level of the track (from 0 to 3). 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 capture switching costs.[39]

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

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

[39]Note that the fixed cost of upgrading is only identified in the first two grades as upgrading afterwards is not feasible and therefore not included in the choice set of students.

Note that in section 2, the scale of the utility function was implicitly normalized to unity. Therefore all parameters μ are identified. However, to directly interpret the cost estimates, I will renormalize the scale by dividing all parameters by μ_{time}. This way, the cost estimates can be measured in daily commuting time.

End-of-year performance

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

In section 3, I explained how a measure of performance can be used to back out the optimal level of effort y_{ijt}^* in a nonparametric way. However, the finite number of observations and the large state space does not allow me to do this. I therefore approximate the optimal level of effort by a parametric structure.[40] The optimal level of effort and the thresholds to obtain each outcome can then be recovered by estimating an ordered logit model with index $\ln(y_{ijt}^*)$. Note that some of the thresholds are not identified from the data but from the institutional context that imposes restrictions on mobility. I also allow the thresholds to differ not only by different programs but also by the grade a student is in.[41]

Note that there is still no need to impose structure on the marginal costs as they are recovered from the marginal benefits at the optimal level of effort (see equation (7) at page 9).

Extension to allow for course-specific restrictions

The model in section 2 only includes one measure of performance. I defined this as the permission to start in each track in the next grade. The problem with this approach is that B-certificates can also exclude elective courses instead of tracks. I therefore extend the model to allow for two additional measures of performance: $g_{ijt+1}^{clas} = \{1, 2\}$ specifies if a students

[40] I impose the same structure on the logarithmic transformation of effort as for the fixed costs, but I also add the effects of distance to higher education institutes and characteristics of last year's program. This is because distance should not have an effect on the fixed cost of schooling in secondary education, but it can have an effect on the optimal level of effort because future utility is affected. I add characteristics of the program a student followed in the previous year to allow experience to affect effort today because of a change in marginal cost of effort. These results can be found in the appendix section A.3.

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

can go to the next grade in a program that includes classical languages. $g_{ijt+1}^{math} = \{1,2,3\}$ specifies if a student can go to the next grade in a math option in the middle-theoretical track ($g_{ijt+1}^{math} = 2$) or the academic track ($g_{ijt+1}^{math} = 3$).[42] I model their distribution by an ordered logit, conditional on the outcome of g_{ijt+1}^{track}, with indexes:

$$g_{ijt+1}^{math\circ} + \eta_{ijt+1}^{math} = \alpha_y^{math} \ln y_{ijt} + S_i' \alpha_S^{math} + \nu_i' \alpha_\nu^{math} + \eta_{ijt+1}^{math} \quad (16)$$

$$g_{ijt+1}^{clas\circ} + \eta_{ijt+1}^{clas} = \alpha_y^{clas} \ln y_{ijt} + S_i' \alpha_S^{clas} + \nu_i' \alpha_\nu^{clas} + \eta_{ijt+1}^{clas}. \quad (17)$$

I also estimate grade- and track-specific thresholds.[43] $\alpha_y^{math} > 0$ and $\alpha_y^{clas} > 0$ measure how much effort, identified from the permissions to start in each track, matters for each elective course. I also allow for comparative advantages in elective courses by estimating the influence of observed and unobserved student characteristics through $(\alpha_S^{math}, \alpha_\nu^{math})$ and $(\alpha_S^{clas}, \alpha_\nu^{clas})$.

5.3 Choice after leaving secondary education

From the age of 18 on, students can leave the education system. If they decide to leave, they can either go to a higher education option ($j \in he$) or choose the outside option $j = 0$ (but might enroll later). If a student leaves secondary education without a degree, he cannot go to higher education. Admission to the first year of higher education is allowed for all students with a high school degree. As explained in section 4.5, I distinguish between three types of higher education, increasing in their academic level: professional college, academic college and universities. In addition, I also distinguish between STEM and non-STEM majors and allow for five campuses to study a university program: Leuven, Ghent, Antwerp, Brussels and Hasselt. For professional and academic colleges, I follow Declercq and Verboven (2017) and assume students choose the closest campus. This results in 15 options after graduating from high school: $j = 0$ or one of the 14 study options $j \in he$.

As explained in section 2.3, the value functions after leaving secondary education can be written as the sum of an estimated, common value of a high school degree μ^{degree} and a choice-specific component. I now impose structure on the choice-specific component $\Psi_j^{HEE}(x_{it}, \nu_i) =$

[42] The downside of this extension is that marginal benefits of effort in the model are not necessarily decreasing over the entire domain of y_{ijt}, making it more difficult to find a solution. Nevertheless, an interior solution is still required and this should satisfy the first-order condition, allowing us to estimate the model as explained in section 3, 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.

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

Ψ_{ij}^{HEE} to let this value differ by the option a student chooses after graduation:

$$\begin{aligned}\Psi_{ij}^{HEE} &= \phi_j^{HEE,0} + S_i'\phi_j^{HEE,S} + \nu_i'\phi_j^{HEE,\nu} + \phi^{HEE,\text{dist}}\text{distance\_HE}_{ij} \\ &+ d_{iT_i^{SE}}'\phi_j^{HEE,SE} + \phi_j^{HEE,delay}\text{delay}_{iT_i^{SE}} \\ &+ \phi_j^{HEE,delay\_level}\text{level\_SE}_{iT_i^{SE}}\text{delay}_{iT_i^{SE}}.\end{aligned} \tag{18}$$

Distance_HE$_{ij}$ is the distance in kilometers from the student's home to the chosen option, $\phi_j^{HEE,SE}$ and $\phi_j^{HEE,delay}$ estimate how secondary education outcomes affect the enrollment decision, with $d_{iT_i^{SE}}$ a vector of dummy variables for each possible program a student can graduate in and delay$_{iT_i^{SE}}$ the years of accumulated study delay. I also include an interaction effect between the academic level, measured by a ranking of the track level_SE$_{iT_i^{SE}}$ and study delay.

Since only differences in utility are identified, I normalize $\phi_0^{HEE} = 0$. By this normalization, all fixed cost parameters in the model should be interpreted as the disutility of going to school, compared to the total value of working (without a degree). Note that this includes effects on labor market outcomes. E.g. if male students experience a higher fixed cost of school than female students, this can be because they dislike school more, relative to work, but also because they might earn higher wages when they choose to work. Because the number of options is large, I need to constrain some of the parameters. I allow for flexible, j-specific constants $\phi_j^{HEE,0}$, but only allow other parameters to differ by the ranking of the academic level of the higher education option (0 to 2) and a dummy that indicates if it is a STEM major.[44]

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

5.4 Additional long run outcomes

I specify three additional long run outcomes: success in the first year of higher education, graduation from higher education before age 25 and the unemployment spell between age 25 and 35. All long run outcomes follow a similar structure as (18) and can be characterized by $w \in W = \{HEE, HES, HED, SPELL\}$.

[44] For this outcome I also do not allow a different effect of each secondary education program on decisions but allow for a general effect of each track on going to college and interactions with the ranking of the academic level of the track, a dummy variable for extra math and a dummy variable for classical languages in the program. Note that this implies that elective courses can only increase enrollment through their effect on programs of higher academic level or STEM-major. This assumption is necessary because very few students in the academic track decide not to enroll in college. I do not need to make this assumption for other long run outcomes.

First year success higher education (HES)

In the first year students can fail, pass without distinction, pass cum laude or pass magna cum laude or higher. I estimate an ordered logit model with an index that is identical to (18) as all variables that are expected to influence enrollment decision are also expected to influence performance:

$$\begin{aligned}\Psi_{ij}^{HES} + \omega_{ij}^{HES} &= \phi_j^{HES,0} + S_i'\phi_j^{HES,S} + \nu_i'\phi_j^{HES,\nu} + \phi^{HES,\text{dist}}\text{distance\_HE}_{ij} \\ &+ d_{iT_i^{SE}}'\phi_j^{HES,SE} + \phi_j^{HES,delay}\text{delay}_{iT_i^{SE}} \\ &+ \phi_j^{HES,delay\_level}\text{level\_SE}_{iT_i^{SE}}\text{delay}_{iT_i^{SE}} \\ &+ \omega_{ij}^{HES}\end{aligned}$$

with ω_{ij}^{HES} logistically distributed. I also allow for interactions of thresholds with the level and major of the program. Similar to the enrollment decision, I allow for flexible j-specific choice-specific constants and constrain other parameters to differ only by the ranking of the academic level of the higher education option and a dummy that indicates if it is a STEM major.

Higher education degree obtained at age 25 (HED)

The model to obtain a higher education degree is similar to the model explaining the decision after leaving secondary education, but adds controls for the enrollment decision and the fact that the student passed the enrollment year. The indexes used to form the logit probabilities are given by:

$$\begin{aligned}\Psi_{ij}^{HED} + \omega_{ij}^{HED} &= \phi_j^{HED,0} + S_i'\phi_j^{HED,S} + \nu_i'\phi_j^{HED,\nu} + \phi^{HED,\text{dist}}\text{distance\_HE}_{ij} \\ &+ d_{iT_i^{SE}}'\phi_j^{HED,SE} + \phi_j^{HED,delay}\text{delay}_{iT_i^{SE}} \\ &+ \phi_j^{HED,delay\_level}\text{level\_SE}_{iT_i^{SE}}\text{delay}_{iT_i^{SE}} \\ &+ Switch_{ij}'\phi^{HED,\text{switch}} + \omega_{ij}^{HED}.\end{aligned}$$

ω_{ij}^{HED} is distributed extreme value type 1 and the vector of a parameters associated to not obtaining a higher education degree are normalized: $\phi_0^{HED} = 0$. $Switch_{ij}$ is a vector of controls for outcomes after leaving secondary education, but before obtaining a degree. This includes a dummy equal to one if the student is in the same higher education level as in the enrollment year (including no higher education as the lowest level), a dummy equal to one if the student is in a higher educational level than in the enrollment year and a dummy equal to one if the student is in the same major as in the enrollment year. Note that this implies that the (full) effect of a study program in secondary education on obtaining a degree at age 25 does not only go through $d_{iT_i^{SE}}$ and $\text{level\_SE}_{iT_i^{SE}}$, but also through $Switch_{ij}'$. All controls are then also interacted with a dummy equal to one if the student passed the enrollment year in higher education. I allow for flexible j-specific choice-specific constants and constrain other parameters to differ only by the ranking of the academic level of the higher education option and a dummy that indicates if it is a STEM major.

Unemployment spell age 25-age 35 ($SPELL$)

Unemployment spell in years is modeled as an ordered logit with thresholds for each year. The index is given by:

$$\begin{aligned}\Psi_{ij}^{SPELL}+\omega_{ij}^{SPELL} \quad = \quad & \phi^{SPELL,0}+S_i'\phi^{SPELL,S}+\nu_i'\phi^{SPELL,\nu} \\ & +d_{iT_i^{SE}}'\phi^{SPELL,SE}+\phi^{SPELL,delay}\text{delay}_{iT_i^{SE}} \\ & +\phi^{SPELL,delay\_level}\text{level\_SE}_{iT_i^{SE}}\text{delay}_{iT_i^{SE}} \\ & +d_{iT_i^{HE}}'\phi^{SPELL,HE}+\omega_{ij}^{SPELL}\end{aligned}$$

with $d_{iT_i^{HE}}$ a vector of indicators for each potential higher education degree obtained at age 25 and ω_{ij}^{SPELL} logistically distributed. The total effect of study programs in secondary education now goes through $d_{iT_i^{SE}}$ and level_SE$_{iT_i^{SE}}$, but also through higher education $d_{iT_i^{HE}}$. The latter is a vector of dummy variables equal to one if the student obtained a specific higher education degree. Note that distance to higher education does not enter the unemployment spell directly, while it does influence graduation from higher education. It therefore serves as an exclusion restriction that helps in identifying the unobserved type distribution.

6 Estimation results

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

6.1 Cost of schooling

In this section I summarize the estimates of fixed costs (C^0 in (15)) and marginal costs of effort c_j. While the structure in the model on the fixed costs is the same as shown in these tables, the marginal costs are derived from the optimal choices of program and effort in the data and are therefore a nonlinear function of the variables and other parameters in the model (see section 3). I therefore perform an OLS regression of the estimated marginal costs with the same structure as the fixed costs to interpret them. All parameters are divided by the parameter of the travel time variable and can therefore be interpreted in terms of daily minutes to travel.[45]

Table 5 shows that high SES students pay a lower fixed costs in the academic track than a student of low SES, equivalent to commuting 282 minutes to school on a daily basis. Also in other programs they pay a lower cost, but to a lesser extent. The impact of SES on marginal cost of effort however is small and insignificant. On the contrary, an increase in math ability by one standard deviation (SD) has no significant impact on fixed costs but decreases the marginal cost of effort by 7.8 minutes. If students would be indifferent between time spent commuting and time spent studying, this would mean that a low math ability student (1 SD lower than the average) can compensate his lower chances of success, compared to the average student, by studying an additional 7.8 minutes per day. We see this trend also in other estimates: SES is mainly affecting the fixed costs, while initial ability is affecting marginal costs of effort. I.e. preferences for going to school and choosing higher level programs are affected mainly by socioeconomic background, but initial ability is more important in explaining the differences in experienced difficulty level. A low ability, high SES student will therefore have to work harder (experience a higher effort costs, due to a larger marginal cost) to achieve the same result. But, since preferences are higher, this student is more willing to pay the additional cost as low effort will be associated with lower chances of staying in the most favored program. This explains why SES has a positive impact on the optimal level of effort in the data and thus on performance, even though there is no effect on the marginal cost of effort.[46]

An important result for the identification approach is the effect of persistent unobserved heterogeneity. The model estimates 67% of students to be of type 1 and 33% to be of type 2 (see appendix Table A1). Type 2 students have lower marginal costs, suggesting they are of higher ability, not captured by the included ability variables. They also have lower fixed costs, especially in more academic programs. This shows that there will be selection on unobservables in the data as type 2 students are more likely to choose more academic programs.

In Table 6, we see that high ability, high SES and type 2 students have lower fixed costs in classical language and math options. For math options, only math and not language ability is relevant. Furthermore, male students experience lower fixed costs in math options. There is no effect on the way marginal costs of effort depend on student background, except for type 2 students who loose the advantage they had in the academic track (classical languages is only available in the academic track). We also see that high SES or belonging to type 2 is

[45] Choice-specific constants can be found in the appendix Table A8.

[46] See appendix section A.3 for the reduced form estimates of the optimal effort level.

Table 5: Costs of schooling: tracks by student background and grade

	Fixed cost		Marginal cost of effort	
	coef	se	coef	se
Academic				
Male	61.0	(60.2)	4.291 **	(1.689)
Language ability	-99.6 **	(44.8)	-7.254 **	(2.951)
Math ability	-42.1	(39.4)	-7.776 ***	(2.550)
High SES	-281.7 **	(125.9)	-1.003	(1.275)
Type 2	-718.3 ***	(169.0)	-3.928 **	(1.587)
Grade	-9.5	(15.5)	0.140	(0.759)
Middle-Theoretical				
Male	-22.3	(58.1)	8.711 ***	(2.714)
Language ability	-20.5	(37.2)	-9.193 **	(3.816)
Math ability	-32.3	(39.2)	-12.431 ***	(3.278)
High SES	-248.4 **	(124.8)	-1.390	(1.848)
Type 2	-678.9 ***	(164.7)	-11.558 ***	(3.022)
Grade	1.2	(13.7)	-0.344	(0.683)
Middle-Practical				
Male	-21.2	(54.7)	7.135 **	(3.094)
Language ability	32.6	(42.2)	-12.647 ***	(4.502)
Math ability	-31.8	(36.9)	-4.433 *	(2.548)
High SES	-239.6 *	(122.4)	-2.331	(3.253)
Type 2	-572.1 ***	(150.6)	-13.758 ***	(4.062)
Grade	-15.4	(12.2)	1.093	(1.257)
Vocational				
Male	24.0	(58.4)	3.136	(2.068)
Language ability	53.5	(44.6)	-8.787 **	(4.445)
Math ability	43.8	(35.7)	-3.076	(2.848)
High SES	-226.0 *	(123.2)	-2.663	(1.818)
Type 2	-497.1 ***	(148.4)	-9.681 ***	(2.690)
Grade	12.8	(13.7)	2.772 ***	(0.957)

Note: Estimates of a sample of 4927 students or 31932 student-year observations during secondary education. Fixed cost estimates of equation (15). The reported marginal costs of effort are an approximation of the predicted values from the model. All parameters are divided by μ_{time} in equation (15) such that they can be interpreted in minutes of daily travel time. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Grade subtracted by 6 to start counting in secondary education. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table 6: Costs of schooling: elective courses, grade interactions and switching costs

	Fixed cost		Marginal cost of effort	
	coef	**se**	**coef**	**se**
Classical languages				
Male	-34.3	(25.9)	-2.319	(1.656)
Language ability	-61.2 **	(25.2)	2.970	(5.097)
Math ability	-71.4 ***	(22.9)	3.268	(6.961)
High SES	-63.0 **	(26.7)	1.294	(1.348)
Type 2	-93.2 ***	(26.2)	4.740 ***	(1.499)
Grade	-24.2 ***	(8.5)	-0.203	(0.602)
Intensive math				
Male	-98.1 ***	(36.2)	-1.930	(1.552)
Language ability	11.9	(27.3)	1.357	(3.695)
Math ability	-167.2 ***	(34.9)	0.685	(3.211)
High SES	-71.7 **	(28.5)	1.528	(1.201)
Type 2	-194.9 ***	(35.5)	0.850	(2.171)
Grade	-59.7 ***	(17.3)	1.761	(1.090)
Grade				
Male	-9.0	(11.2)	0.533	(0.515)
Language ability	-8.5	(9.3)	0.278	(0.906)
Math ability	-7.2	(6.8)	0.317	(0.618)
High SES	35.8 *	(20.9)	-0.742 *	(0.385)
Type 2	88.7 ***	(27.2)	-1.736 ***	(0.551)
Study delay	87.7	(58.6)	8.110	(5.703)
Grade	-15.4	(11.1)	0.285	(1.213)
Level SE	8.4	(9.9)	-0.322	(1.368)
Repeat grade	387.0 ***	(120.5)	-14.087 ***	(5.406)
Grade	4.1	(18.9)	4.119 **	(1.774)
Level SE	138.8 ***	(28.7)	-3.090 **	(1.466)
Downgrade	270.3 ***	(46.1)	0.213	(0.961)
Upgrade	545.6 ***	(106.9)	2.954	(2.502)
Time	1.0		-0.003	(0.004)

Note: Estimates of a sample of 4927 students or 31932 student-year observations during secondary education. Fixed cost estimates of equation (15). The reported marginal costs of effort are an approximation of the predicted values from the model. All parameters are divided by μ_{time} in equation (15) such that they can be interpreted in minutes of daily travel time. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Grade subtracted by 6 to start counting in secondary education. Level SE = academic level of study program. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

more important for fixed costs in lower grades but shifts to marginal cost in higher grades.

We also see the importance of variables that are influenced by the tracking policy. We do not find a significant effect of study delay, i.e. past grade retention, but a strong increase in fixed costs of currently repeating a grade, especially in programs of higher academic level. At the same time, repeating the grade also decreases marginal costs, especially in higher level programs. This shows a clear trade-off: students dislike repeating a grade but it does help them to perform well, especially in more academic programs. This is one of the explanations why students might consider repeating a grade, even if they have the possibility to go to the next grade in another program. Students also 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.

6.2 Long run outcomes

All estimates for long run outcomes can be found in appendix section A.4. I hereby discuss the main results.

Students value the long run impact of their program. First, I find positive estimates of the value of a degree. Furthermore, students obtain an additional benefit of being allowed to enroll in higher education. Male students are less likely to have better higher education outcomes, but they do have smaller unemployment spells. Initial ability matters especially for higher education programs of higher academic level. Math ability is particularly important for enrollment in and graduation from STEM programs and also decreases unemployment spells. High SES students have better higher education outcomes but do not experience an effect on unemployment. For the identification strategy, it also important to point out that unobserved type is again important. While type 2 students are not more likely to enroll in professional colleges, they are more likely to choose a STEM major and institutions of higher academic level. Moreover they perform better and are more likely to obtain a higher education degree. We can therefore conclude that type 2 is of higher ability, according to long run outcomes. Note that this type also has lower schooling costs in secondary education, especially in programs of higher academic level. This means that there will be self-selection into programs by high ability students into programs that showed better outcomes in a descriptive analysis. Therefore, it is important to control for type to avoid ability bias in the estimates and counterfactuals.

Because the estimates of high school programs on long run outcomes are difficult to interpret, I also calculate the total Average Treatment Effects on the Treated (ATT) of each study program in appendix section A.5. These estimates look at the total causal impact of the high school program students graduates in on long run outcomes, after conditioning on other high school outcomes like study delay and drop out. Most estimates point in the same direction as a simple comparison of means in the data, but to a smaller extent. I find that graduating from the academic track without classical languages or extra math leads to an increase in college graduation of 25 %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 (12 %points). Elective courses mainly matter for type of higher education. Students who choose classical languages are 20 %points more likely to graduate from universities, but 19 %points less likely to graduate from professional colleges. Extra math in the academic track leads to an increase of 12 %points in STEM degrees and extra math in the middle-theoretical track to an increase of 19 %points.

Tracking policies alter the trade-off between the academic level of the program a student is in and years of study delay he accumulates. It is therefore also interesting to look at the ATT of study delay. I compare the effect for students with zero and one year of study delay for those that are retained during secondary education. I find a negative effect on higher education outcomes, but only statistically significant for the probability to obtain a higher education degree. One year of study delay decreases the probability to obtain a higher education degree by 10 %points. Also the unemployment spell is affected as it increases by 3.6 months on a time span of 10 years.

6.3 Model fit

After estimation, I solve the model as explained in section 2.3. Once I have solved the model backwards to find all the conditional value functions and effort levels, I forward simulate all error terms to simulate choices and performance and let the model generate a database of predictions. I then compare this database to the actual data and use it to compare it to counterfactual scenarios. To allow students to change their effort level, I perform a grid search over different effort levels in each conditional value function to look for the optimal value.[47] Table 7 shows the ability of the model to replicate the actual data if there is no policy change. The model does a decent job in predicting the graduation track, although there is a slight overprediction of graduating from the academic track. The main outcomes

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

Table 7: Predictions of the model

		Data	Predictions
High school graduation (% of students)			
Academic		*36.3*	*43.3 (2.9)*
	clas+math	3.7	4.7 (1.0)
	clas	5.5	5.1 (1.1)
	math	12.6	15.6 (1.5)
	other	14.4	17.9 (1.8)
Middle-Theoretical		*16.4*	*11.1 (2.1)*
	math	2.6	2.1 (0.6)
	other	13.8	9.0 (1.5)
Middle-Practical		*12.7*	*11.2 (2.1)*
Vocational		*20.5*	*21.3 (1.0)*
Dropout		*14.2*	*13.2 (0.5)*
Students with grade retention		33.2	32.0 (1.2)
Higher education (% of students)			
Enrollment		55.2	56.8 (1.0)
First year successful (among enrolled)		45.6	45.1 (1.2)
Degree (age 25)		39.0	40.2 (0.8)
University degree		9.6	8.6 (0.7)
Academic college degree		4.8	5.0 (0.4)
Professional college degree		24.8	26.7 (0.9)
Degree in STEM major		10.2	9.8 (0.6)
Unemployment (mean in years)			
Spell age 25-35		2.08	2.03 (0.05)

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

of interest are predicted very precisely. This holds for outcomes in secondary education like grade retention and drop out, but also outcomes in higher education and the average unemployment spell.

7 Added value of high school programs

In the previous section I interpreted the estimates on long run outcomes to look at the effect of high school programs for individual students. These estimates are "ceteris paribus" causal effects, i.e. they are the effect of one variable if all other variables that were realized before leaving secondary education are kept fixed. The ceteris paribus effects are problematic when we think about policy recommendations. This is because enabling students to choose a certain program can also influence other outcomes during secondary education like preferences to go to school and the willingness to exert effort. This will influence other endogenous variables like drop out and grade retention. I therefore present an alternative way to assess the causal impact of high school programs. I investigate their added values by simulating choices in a world where one track or elective course is not available and look at the way it influences outcomes of students. The added values I report are differences between the status quo, i.e. all options are available, and a world where some programs are omitted. In this section I only discuss the added value estimates. See appendix section A.6 for the results of these simulations.

I first discuss the added value estimates, and then compare my results to a traditional dynamic model where effort is assumed to be exogenous.

7.1 Added value estimates

Table 8 shows the estimates of the added value of each track and elective course on the main higher education outcomes and the unemployment spell. The impact of high school programs is similar for the three higher education outcomes. The academic track is responsible for an increase of 8.6 %points in graduation from higher education. For the vocational track, we observe the opposite effect. It decreases graduation by 1.9 %points. However, without the vocational track, unemployment spells would increase. Elective courses also matter for long run outcomes. Classical languages increase graduation from college by 1.8 %points. Extra math increases enrollment but the effect on graduation from college is small and insignificant.

Table 9 discusses the effect on different options of higher education in more detail. While elective courses within the academic track only had small effects on general outcomes, they are particularly important for the academic level and major of the higher education programs

Table 8: High school program and long run outcomes: added value

	Higher education (added value in %points)			**Unemployment** (added value to mean)
	Enrollment (% of students)	First year success (% of enrolled)	Degree age 25 (% of students)	Spell age 25-35 (mean in years)
Tracks				
Academic	+8.53 *** (1.19)	+3.21 (2.83)	+8.64 *** (1.37)	-0.02 (0.05)
Middle-Theoretical	+0.28 (0.43)	-0.73 (0.52)	-0.62 (0.39)	+0.01 (0.01)
Middle-Practical	-0.27 (0.33)	-0.42 (0.36)	-0.57 ** (0.28)	-0.03 * (0.02)
Vocational	-3.72 *** (0.46)	+0.75 *** (0.29)	-1.90 *** (0.30)	-0.09 ** (0.04)
Elective courses				
Classical languages	+1.82 *** (0.24)	+0.60 (0.58)	+1.85 *** (0.38)	-0.03 ** (0.01)
Intensive math	+1.14 *** (0.25)	-1.39 (1.41)	+0.18 (0.66)	-0.01 (0.03)
Predicted value	56.8	45.1	40.2	2.0
Data	55.2	45.6	39.0	2.1

Note: Added values calculated by comparing predictions from a world without the track or elective course with the status quo. Differences in percentage points (higher education) or years (unemployment). Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

in which students obtain a degree. The added value of classical languages on obtaining a higher education degree is almost exclusively explained by an increase in graduation from the most prestigious higher education option: universities. Math-intensive programs did not lead to a significant increase in college graduation but do explain a shift to STEM majors. Without math-intensive programs in high school, the number of students with a STEM degree would decrease by 2.5 %points, a high number given that only 10% of students obtain a STEM degree.

Finally, Table 10 shows the impact of the availability of tracks and elective courses on outcomes during secondary education: grade retention and drop out. The vocational track is very important in avoiding grade retention and drop out. Without the vocational track, students of lower academic ability would have been forced to study a more academically rigorous program if they wanted to graduate with a high school degree. The required increase in effort leads many students to drop out or to accumulate study delay in order to graduate. Surprisingly, we also see that the academic track decreases grade retention and drop out,

Table 9: High school program and level and major college degree: added value in %points

	Academic level higher education						Major	
	University (% of students)		Academic college (% of students)		Professional college (% of students)		STEM (% of students)	
Tracks								
Academic	+4.68 ***	(0.62)	+1.21 ***	(0.25)	+2.75 **	(1.14)	-1.31	(0.81)
Middle-Theoretical	-0.34 ***	(0.09)	-0.16 ***	(0.05)	-0.12	(0.36)	+0.24 *	(0.14)
Middle-Practical	-0.10 ***	(0.03)	-0.07 **	(0.03)	-0.40	(0.25)	+0.20 **	(0.09)
Vocational	-0.02 **	(0.01)	-0.03 **	(0.01)	-1.85 ***	(0.29)	-0.15 **	(0.07)
Elective courses								
Classical languages	+1.92 ***	(0.44)	+0.16 *	(0.09)	-0.23	(0.36)	-0.40	(0.24)
Intensive math	-0.39	(0.65)	+0.47 ***	(0.17)	+0.10	(0.67)	+2.49 ***	(0.75)
Predicted value	8.57		4.95		26.68		9.81	
Data	9.58		4.76		24.85		10.21	

Note: Added values calculated by comparing predictions from a world without the track or elective course with the status quo. Differences in percentage points (higher education) or years (unemployment). Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

albeit to a lesser extent. This can be explained by effort choices. I illustrate this by comparing the prediction of the model to the prediction of a model where effort is not modeled as a choice.

Table 10: High school program and grade retention and drop out: added value in %points

	Grade retention (% of students)	**Drop out** (% of students)
Tracks		
Academic	-3.10 ** (1.45)	-0.68 *** (0.25)
Middle-Theoretical	+2.70 *** (0.39)	+0.91 *** (0.21)
Middle-Practical	+1.90 *** (0.38)	+0.46 (0.37)
Vocational	-9.40 *** (0.68)	-10.72 *** (1.03)
Elective courses		
Classical languages	-2.90 *** (0.62)	-0.31 *** (0.11)
Intensive math	+2.10 *** (0.81)	+0.09 (0.11)
Predicted value	32.0	13.2
Data	33.2	14.2

Note: Added values calculated by comparing predictions from a world without the track or elective course with the status quo. Differences in percentage points (higher education) or years (unemployment). Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

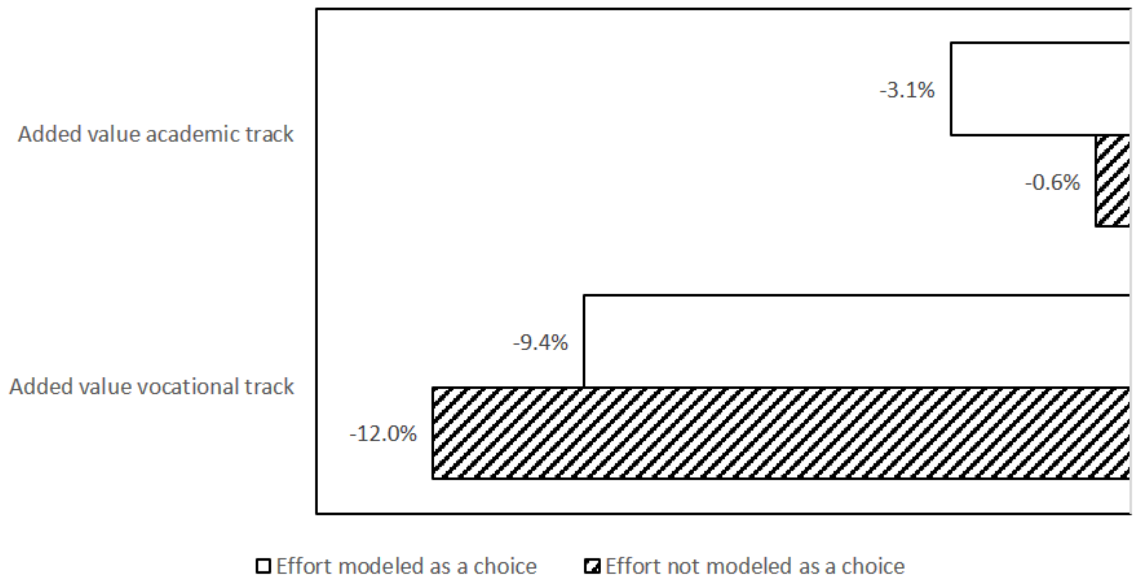
7.2 Bias in added values if effort choice is ignored

In the Appendix Table A22, I compare the results from counterfactual simulations to the results of the traditional model that assumes effort is exogenous. I do this by estimating the flow utility (equation (5)) with marginal costs $c_j = 0$, and by not updating the optimal level of effort we see in the data. While the predictions from the model in the status quo scenario remain very similar, the counterfactual predictions show some important differences in magnitude.[48] To illustrate this, Figure 2 compares the added value on grade retention of the academic track and the vocational track.

There are strong biases in the estimated effects. While the decrease in grade retention because of the academic track is underestimated without effort choice, it is overestimated for the vocational track. To explain this difference, it is important to see how effort responds to each counterfactual. High effort will make it more likely to be admitted to the academic track in the next grade. If the academic track is not available, it makes the benefits smaller,

[48] Predictions of the model without effort are available upon request.

Figure 2: Impact of modeling effort on added value estimates academic and vocational track for grade retention



Note: Added value estimates for the percentage of students with grade retention. Differences in percentage points. Model without effort choice uses estimates of a model with marginal costs = 0 and the effort level in the data. Differences between two models are statistically significant, see appendix Table A22 and Table A23.

making students set lower levels of effort. By setting lower levels of effort, students also become more likely (need smaller negative performance shocks) to not be admitted to any program in the next grade, requiring them to repeat the grade. A model that ignores effort choice will assume students set the same levels of effort as before in a given program, making the increase in grade retention in a world without the academic track much smaller. If the vocational track is removed, we see the opposite effect. To proceed to the next grade, students can count on the vocational track if their effort is too low for other tracks. However, without this possibility, they have to put in enough effort to reach at least the middle-practical track. The benefits of effort therefore increase, students set higher levels of effort and become less likely to be retained. The vocational track is still responsible for a large decrease in grade retention, but not as much as a model without effort would predict.

Many biases in other outcomes can be explained as a result of the bias in predicted grade retention. In the appendix section A.6, I show all the biases of the simulations.

8 Tracking policies in secondary education

In the current tracking policy in Flanders, teachers decide if a student has acquired the necessary skills to transition to the next grade in each of the programs. In some cases, students have not acquired the skills to transition to the next grade, regardless of their program choice. They then obtain a C-certificate which requires them to repeat the grade. However, in many cases students are allowed to transition to the next grade, but have to switch to a program of lower academic level.[49] In this case they obtain a B-certificate. This allows underperforming students to avoid grade retention, however they can still opt for the same program if they are willing to repeat the grade. In a first counterfactual, I look at the effect of this policy by removing the option to avoid grade retention and force students to repeat the grade if they underperformed during this year. In a second counterfactual, I instead reinforce this policy of avoiding grade retention, by forcing them to downgrade instead of having a choice between repeating the grade or downgrading.

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

8.1 Changes in the B-certificate policy

In Table 11 I compare the outcomes of the two counterfactuals to the status quo. Appendix Table A27 shows how the number of students who graduate in each program changes. The policy "Repeat" forces students to repeat the grade after obtaining a B-certificate, i.e. removing downgrading as a way to avoid grade retention. The policy "Downgrade" forces students to downgrade after a B-certificate by not allowing them to repeat the grade.

Without the ability to avoid grade retention, outcomes would have been worse. Although the policy increases graduation from the academic track slightly, it comes at the cost of an increase in students with grade retention by 9.5 %points and increase in drop out rates by 4.0 %points. Also after secondary education, we only see negative effects: enrollment in higher education decreases by 1.9 %points and graduation by 1.8 %points. The average unemployment spell increases by 0.12 years or 6%. The current policy is therefore better than a strict pass or fail policy. Nevertheless, it can be further improved.

If students who obtained a B-certificate were not allowed to repeat the grade, grade retention would decrease by 9.6 %points and drop out by 1.5 %points. This does come at a cost in the short run. Students switch to programs of lower academic level, which decreases enrollment in higher education by 1.2 %points. However, this policy only decreases

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

Table 11: Predictions of the model: counterfactual tracking policy

	Status quo	**Policy change B-certificate**	
		Repeat	Downgrade
High school graduation (% of students and change in %points)			
Students with grade retention	32.0	+9.50 *** (0.57)	-9.60 *** (0.72)
Dropout	13.2	+4.03 *** (0.35)	-1.45 *** (0.18)
Higher education (% of students and change in %points)			
Enrollment	56.8	-1.90 *** (0.34)	-1.21 *** (0.18)
First year successful (among enrolled)	45.1	+0.11 (0.22)	+0.81 ** (0.33)
Degree (age 25)	40.2	-1.75 *** (0.32)	+0.32 (0.22)
University degree	8.6	-0.13 ** (0.07)	+0.30 *** (0.09)
Academic college degree	5.0	-0.16 *** (0.04)	+0.09 * (0.05)
Professional college degree	26.7	-1.46 *** (0.26)	-0.07 (0.18)
Degree in STEM major	9.8	-0.48 *** (0.10)	+0.27 * (0.16)
Unemployment (mean in years and difference in means)			
Spell age 25-35	2.03	+0.12 *** (0.02)	-0.05 *** (0.01)

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

enrollment for students with low chances of eventually graduating from college. The first year success rate increases by 0.8 %points and the number of students who graduate with a higher education degree even slightly increases, albeit not significantly. Also unemployment spells decrease slightly.

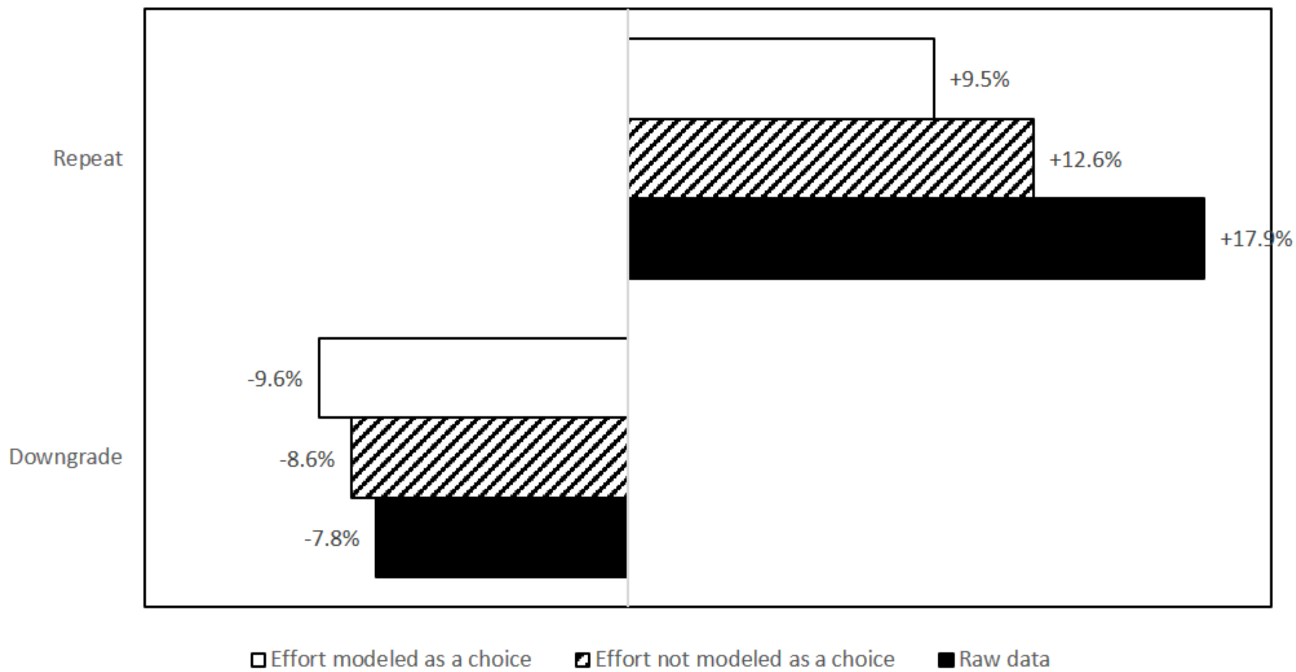
I conclude that the current tracking policy is a good alternative to guide students in their track choices, rather than having them repeat a grade if they fail. Nevertheless, the choice they currently have to repeat a grade instead of downgrading does not lead to beneficial effects in the long run. Given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

8.2 Bias in effects tracking policy if effort choice is ignored

I compare these results again with predictions from a model that ignores effort is a choice by estimating the flow utility with marginal costs $c_j = 0$, and by not updating the optimal level of effort we see in the data. Appendix Table A28 contains the biases of predicted outcomes and Figure 3 illustrates this for the effect on predicted grade retention.

I compare the model with and without effort choice, but also a model-free naive estimate. The latter is calculated by counting the students who currently choose (not) to repeat a grade after obtaining a B-certificate. For the "Repeat" counterfactual, this is the percentage of students who were not retained during secondary education because they chose to downgrade after obtaining a B-certificate. For the "Downgrade" counterfactual, this is the percentage of students who obtained grade retention because they chose to repeat a grade after obtaining a B-certificate. The naive estimate clearly overpredicts the increase in grade retention when students have to repeat the grade, and underpredicts the decrease if they have to downgrade, compared to dynamic choice models. A similar conclusion can be made by comparing the model with and without effort choice in which the model with effort choice leads to more favorable outcomes. The reason why we see these differences is because each model accounts for behavioral responses by students in a different way. Both counterfactuals give students less options to choose from when obtaining a B-certificate. Therefore, they have a reason to avoid obtaining one. A naive estimate from the data ignores that students know about the new policy and will make choices to avoid a B-certificate. It also ignores that students who escape from grade retention, might be more likely to be retained in the future anyway. In a model with dynamic program choice, students can escape from having a B-certificate by choosing another (easier) program. The model also accounts for the fact that this student might get different results in the future. In a model that includes effort choice, students have

Figure 3: Impact of modeling effort on change in grade retention because of change in tracking policy



Note: Estimated effects of counterfactual tracking policies on the percentage of students with grade retention. Differences in percentage points. 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. Model without effort choice uses estimates of a model with marginal costs = 0 and the effort level in the data. Predicted effects from the raw data for the "Repeat" counterfactual: percentage of students who were not retained during secondary education because they chose to downgrade after obtaining a B-certificate. Predicted effects from the raw data for the "Downgrade" counterfactual: percentage of students who obtained grade retention because they chose to repeat a grade after obtaining a B-certificate in the "Downgrade" counterfactual. Differences between the models are statistically significant, see appendix Table A28.

an additional channel to avoid obtaining a B-certificate by increasing their effort. Since they particularly dislike repeating a grade, the differences are largest for the policy in which they are not allowed to downgrade to avoid grade retention.

9 Limitations

As any model is a simplification of reality, there are some limitations that are important to acknowledge. As most of the literature on the returns to education, I abstract from general equilibrium effects. One form of general equilibrium effects is peer effects. The estimated effects of study programs can capture the effects of peers. Counterfactual simulations then ignore the impact of a change in the peer composition in the classroom. Since the counterfactual tracking policies did not cause a large shift in graduation from each program, this is unlikely to be of first order importance. However, future research that wants to address the effect of tracking itself can benefit from modeling peer effects. A second source of general equilibrium effects lies in the long run outcomes. In the current institutional context, this is of minor importance for higher education as there is no competition for spots due to the lack of admission standards. Nevertheless, the percentage of students who are successful and eventually graduates can be the result of preferences by higher education institutes that aim at certain numbers of college graduates. Also unemployment spells can be affected by others' behavior. If extra students on the labor market crowd out others, the effects in the model are overestimated. The unemployment spell should therefore be seen as a proxy for labor market outcomes by looking how study programs and counterfactuals change the willingness and possibility to find a job, through changes in labor-leisure preferences, productivity or signaling.

A second concern is the assumption on the information of students. There is a growing literature on the importance of imperfect information about own ability. Some models therefore allow for students to learn about their ability (see e.g. Stinebrickner and Stinebrickner (2014) and Arcidiacono *et al.* (2016)), instead of making the standard assumption that they know it from the start. If there is in fact learning, I do expect fixed costs of higher tracks to be underestimated because they also help more in giving information about the academic ability of the student. However, since the counterfactual simulations are unlikely to change this learning benefit, I expect this to be a minor issue for the analysis in this paper. Nevertheless, it could be interesting to investigate the effects of policies that reveal more information about ability to students to improve the match between student and program. Future research can therefore look at how to combine a model of endogenous effort with learning about ability.

Finally, I do not look at the specific impact of each school but model the choice of a study program only. It is therefore important to stress that the effect of a program can be the effect of the curriculum but also the quality of the schools where the study program is available. Schools can also differ in the way they set standards to go to the next grade in each program. Since I do not model how schools and teacher decide to set standards, I keep them fixed in counterfactual simulations. It would be interesting to investigate if the standards itself can be set more optimally, rather than changing the implications of B-certificates.

10 Conclusion

I estimated a dynamic choice model of program and effort choices in secondary education to identify the causal effect of high school programs on long run educational outcomes and on unemployment spells. Using a dataset for Flanders (Belgium), I find that academic programs are important for educational attainment, while vocational programs keep students in schools and therefore decrease long run unemployment spells. Nevertheless, policies that encourage students who underperform to opt for programs of lower academic level do not have a negative effect on long run outcomes and significantly decrease grade retention and drop out from high school. This shows that small changes to tracking policies in secondary education can have important effects. Future research should therefore focus more on how tracking policies can improve student outcomes.

From a methodological perspective, I show that it is important to control for the fact that students can change their effort levels in response to counterfactual policy changes. In general, effort allows students to avoid some of the negative consequences of imposed policies. This is especially the case when effort is expected to increase due to more strict policies. This turned out to be important in the current analysis, especially because of its effect on predicted grade retention. Further research can apply the modeling strategy to other contexts where dynamics are important. Not only models of human capital accumulation can benefit from this approach. This innovation also applies to models of investment in physical capital as state transitions are often under the control of the decision maker. The depreciation of capital can then be allowed to depend on the intensity of capital use, which is often not exogenous but one of the decision variables of a firm.[50]

[50] E.g. in the seminal paper by Rust (1987), agents decide on when to replace bus engines, depending on current and expected future mileage and engine prices. While mileage itself is treated as stochastic, it is reasonable to assume that its predictable component is controllable and an increase in future prices of bus engines can have an impact on the miles busses are expected to drive.

References

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

Altonji J G, Arcidiacono P & Maurel A (2016). The analysis of field choice in college and graduate school: Determinants and wage effects, *in Handbook of the Economics of Education*, 5, pp. 305–396.

Altonji J G, Blom E & 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 M & Hotz V J (2016). University Differences in the Graduation of Minorities in STEM Fields: Evidence from California, *American Economic Review* **106**(3), 525–562.

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

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

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

Arellano M & Bonhomme S (2017). Nonlinear panel data methods for dynamic heterogeneous agent models, *Annual Review of Economics* (0).

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

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

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

Cockx B, Dejemeppe M, Launov A & Van der Linden B (2017). Imperfect monitoring of job search: Structural estimation and policy design, *Journal of Labor Economics* **forthcoming**.

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

De Groote O & Declercq K (2017). Tracking and specialization of high schools: heterogeneous effects of school choice, *Working paper* .

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

Declercq K & Verboven F (2017). Enrollment and Degree Completion in Higher education without Admission Standards, *Working paper* .

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

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

Eurydice European Unit (2011). *Grade retention during compulsory education in Europe: regulations and statistics*, Education, Audiovisual and Culture Executive Agency, Brussels. OCLC: 901698490.

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

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

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

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

Joensen J S (2009). Academic and labor market success: The impact of student employment, abilities, and preferences, *Manuscript, Stockholm School of Economics* .

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

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

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

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

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

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

OECD (2013). *PISA 2012 Results: What Makes Schools Successful (Volume IV)*, PISA, OECD Publishing. DOI: 10.1787/9789264201156-en.

OECD (2017). *Education at a Glance 2017*, Education at a Glance, OECD Publishing. DOI: 10.1787/eag-2017-en.

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 & 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 & 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.

Stinebrickner R & Stinebrickner T R (2014). A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout, *The Review of Economic Studies* **81**(1), 426–472.

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

A Appendix

A.1 Curvature of marginal benefits

This section discusses the shape of the marginal benefits of effort that were introduced in section 2. Note that marginal costs of effort are assumed to be constant and effort lies in the open interval $(0, +\infty)$. The marginal benefits should therefore be sufficiently flexible to guarantee an interior solution. I show that the marginal benefits are positive, decreasing in effort and follow an S-shaped curve.[51]

Marginal benefits are positive

In the paper section 2, I described that the marginal benefits of effort are given by

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

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

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

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

Marginal benefits are decreasing in effort

First note that $\beta\left(\bar{V}_{t+1}(x_{it+1}(g_{it+1}+1), \nu_i) - \bar{V}_{t+1}(x_{it+1}(g_{it+1}), \nu_i)\right)$ is always positive and does not depend on y_{ijt}. Therefore, a sufficient condition for the marginal benefits to be

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

decreasing is $\frac{\partial \frac{(1-P_j(g_{it+1}\leq \bar{g}|y_{ijt}))P_j(g_{it+1}\leq \bar{g}|y_{ijt})}{y_{ijt}}}{\partial y_{ijt}} < 0 \; \forall g_{it+1} < G$:

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

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

Marginal benefits are S-shaped

Note that we can rewrite

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

Marginal benefits then become

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

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

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

$P_j(g_{it+1}\leq \bar{g}|y_{ijt}) \to$constant if $y_{ijt}\to +\infty$ (with the constant 0 for all probabilities, except for the largest performance measure where the limit is 1). Therefore the upper limit is:

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

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

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

A.2 CCP estimation without terminal action

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

The choice probabilities (9) can also be written by using differenced value functions:

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

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

for any $j^{'} \in \Phi(x_{it})$ and $u_j(x_{it}, \nu_i) = -C_j^0(x_{it}, \nu_i) - c_j(x_{it}, \nu_i) y_{ijt}^*$, with $y_{ijt}^* = y_{jt}^*(x_{it}, \nu_i)$. Substitute the CCP representation of the future value as a function of the CCP of an arbitrary choice and its conditional value function (12) in (19):

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

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

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

with $P_{d^*}(g_{i\tau+1} = \bar{g} | y_{d^*t}^*(x_{i\tau}, \nu_i))$ the probability of receiving performance outcome $g_{i\tau+1} = \bar{g}$ at time $t = \tau$, in the program a student will be according to the decision sequence d_i^*. Similarly, define $\kappa_\tau^{'}$ to be the transitions in a sequence where the choice in t is different: $d_i^{'} = (d_{it}^{'}, d_{it+1}^*, d_{it+2}^*, ...d_{it+\rho_{it}}^*)$.[52] We can then repeat the CCP method in each of the future periods and rewrite (20) as the sum of future flow utilities and CCPs until the outside option becomes available at $t + \rho_{it}$:

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

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

$\overline{V}_{t+\rho_{it}}$, the value of behaving optimally when the outside option is available, can be written as in (13). 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. However, further simplifications follow from a good choice of the arbitrary options and a convenient parameterization of the model.

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

The same procedure is applied within $u_j(x_{it}, \nu_i) = -C_j^0(x_{it}, \nu_i) - c_j(x_{it}, \nu_i) y_{jt}^*(x_{it}, \nu_i)$. By substituting the marginal cost of effort by the marginal benefit of effort, future value terms also enter directly into $u_j(x_{it}, \nu_i)$ (see (11)). Because $\sum_{\bar{g}} \frac{\partial P_j(\bar{g}|y_{ijt})}{\partial y_{ijt}} = 0$, all terms that do not depend on performance drop out such that the same simplifications arise because of finite dependence.

A.3 Additional estimation results secondary education

Table A1: Type probabilities in %

	Type probabilities	
	Type 1	Type 2
Overall	*67.3*	*32.7*
Age 12	63.5	36.5
Age 13	88.9	11.1
Age 14	87.8	12.2

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

Table A2: Estimates of optimal effort level (1)

	Log of optimal effort		
	coef		se
Academic			
Male	-0.561	***	(0.130)
Language ability	0.755	***	(0.213)
Math ability	1.037	***	(0.116)
High SES	0.439	***	(0.149)
Type 2	1.069	***	(0.165)
Grade	-0.009		(0.046)
Middle-Theoretical			
Male	-0.551	***	(0.133)
Language ability	0.852	***	(0.132)
Math ability	0.865	***	(0.098)
High SES	0.255	*	(0.148)
Type 2	0.980	***	(0.163)
Grade	-0.056		(0.037)
Middle-Practical			
Male	-0.314	**	(0.128)
Language ability	0.812	***	(0.123)
Math ability	0.425	***	(0.105)
High SES	0.232		(0.202)
Type 2	0.712	***	(0.177)
Grade	-0.062		(0.046)
Vocational			
Male	-0.326	**	(0.147)
Language ability	0.655	***	(0.141)
Math ability	0.352	***	(0.112)
High SES	0.500	**	(0.250)
Type 2	1.371	***	(0.277)
Grade	-0.202	***	(0.052)

Note: Estimates of approximation of the optimal level of effort in the data. Clas= classical languages included. Math= intensive math. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Grade subtracted by 6 to start counting in secondary education. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A3: Estimates of optimal effort level (2)

	Log of optimal effort		
	coef		**se**
Academic			
clas+math	-0.925		(0.772)
clas	-0.683		(0.451)
math	-1.073	**	(0.487)
other	-1.162	***	(0.322)
Middle-Theoretical			
math	-0.825	*	(0.492)
other	-1.047	***	(0.292)
Middle-Practical	-1.174	***	(0.250)
Vocational			
Part-time			
Grade			
Male	-0.026		(0.028)
Language ability	-0.075	***	(0.028)
Math ability	-0.049	**	(0.023)
High SES	0.010		(0.033)
Type 2	0.081	**	(0.034)
Time	0.000		(0.001)
Constant	3.523	***	(0.290)

Note: Estimates of approximation of the optimal level of effort in the data. Clas= classical languages included. Math= intensive math. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Grade subtracted by 6 to start counting in secondary education. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A4: Estimates of optimal effort level (3)

	Log of optimal effort	
	coef	**se**
Classical languages		
Male	-0.564 **	(0.287)
Language ability	0.958 **	(0.376)
Math ability	0.065	(0.390)
High SES	0.325	(0.360)
Type 2	0.848 **	(0.371)
Grade	-0.006	(0.127)
Intensive math		
Male	-0.225	(0.172)
Language ability	0.407 **	(0.205)
Math ability	0.096	(0.205)
High SES	-0.032	(0.201)
Type 2	0.560 ***	(0.213)
Grade	-0.121	(0.087)
Study delay	-0.505 ***	(0.186)
Grade	0.352 ***	(0.112)
Level SE	-0.008	(0.049)
Repeat grade	0.743 **	(0.289)
Grade	0.352 ***	(0.112)
Level SE	0.272 ***	(0.092)
Downgrade	0.144 **	(0.073)
Upgrade	0.074	(0.120)

Note: Estimates of approximation of the optimal level of effort in the data. Clas= classical languages included. Math= intensive math. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Grade subtracted by 6 to start counting in secondary education. Level SE = academic level of study program. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A5: Estimates of optimal effort level (4)

	Log of optimal effort	
	coef	**se**
Last year academic level secondary education	-0.012	(0.037)
Last year classical languages	0.186	(0.124)
Last year intensive math	-0.367 ***	(0.142)
Distance professional college - STEM	-0.006	(0.011)
x academic level secondary education	0.000	(0.006)
Distance professional college - No STEM	0.009	(0.010)
x academic level secondary education	0.001	(0.004)
Distance academic college - STEM	0.008	(0.010)
x academic level secondary education	-0.002	(0.005)
Distance academic college - No STEM	0.028	(0.017)
x academic level secondary education	-0.022 **	(0.010)
Distance university	-0.024	(0.022)
x academic level secondary education	0.022 *	(0.012)

Note: Estimates of approximation of the optimal level of effort in the data. Clas= classical languages included. Math= intensive math. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Grade subtracted by 6 to start counting in secondary education. Bootstrap standard errors between parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$ (normal-based).

Table A6: Estimates of optimal effort level (5)

	Log of optimal effort		
	coef		se
Constant	3.523	***	(0.290)
Threshold level 2	0.000		(.)
Threshold level 3	0.804	***	(0.046)
Threshold level 4	0.870	***	(0.048)
Threshold level 5	0.919	***	(0.051)
Academic			
x Threshold level 3	-0.554	***	(0.047)
x Threshold level 4	-0.391	***	(0.062)
x Threshold level 5	-0.135	*	(0.080)
Grade 8			
x Threshold level 3	0.000		(.)
x Threshold level 4	0.440	***	(0.045)
x Threshold level 5	0.869	***	(0.055)
Grade 9			
x Threshold level 3	-0.142	***	(0.049)
x Threshold level 4	0.118	*	(0.066)
x Threshold level 5	0.368	***	(0.077)
Grade 10			
x Threshold level 3	-0.138	***	(0.051)
x Threshold level 4	0.229	***	(0.070)
x Threshold level 5	0.544	***	(0.081)

Note: Threshold level 2 is normalized to 0. Grade 8 x Threshold level 3 is also set to 0 because outcome 1 impossible in first grade. From grade 11 on, switching programs is no longer possible in the next grade so thresholds are not estimated but become either 0 or infinity. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A7: Estimates of impact effort and comparative advantage on clas and math certificates

	Classical languages coef	se	**Intensive math** coef	se
Log of effort	0.473	(0.446)	0.347 **	(0.170)
Male	0.446	(0.681)	-0.430 *	(0.225)
Language ability	0.270	(0.772)	0.080	(0.217)
Math ability	-0.225	(0.642)	0.023	(0.228)
High SES	-0.071	(0.503)	-0.233	(0.258)
Type 2	0.379	(1.044)	0.215	(0.431)
Threshold level 2	-0.365	(1.623)		
Threshold level 2 x grade	-0.044	(0.180)		
Academic				
x Threshold level 2			-2.275 ***	(0.309)
x Threshold level 3			-1.385 ***	(0.310)
Grade 8				
x Threshold level 2			-2.239 ***	(0.445)
x Threshold level 3			-2.083 ***	(0.440)
Grade 9				
x Threshold level 2			-2.074 ***	(0.461)
x Threshold level 3 (academic only)			-2.339	(3.481)
Grade 10				
x Threshold level 2			-1.796 ***	(0.452)
x Threshold level 3 (academic only)			-0.632	(0.443)

Note: Estimates of equations (16) and (17). 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. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A8: Costs of schooling: choice-specific constants

	Fixed cost		Marginal cost of effort	
	coef	**se**	**coef**	**se**
Academic				
clas+math	770.0 ***	(145.9)	1.837	(6.579)
clas	259.5 ***	(72.0)	8.230 ***	(2.741)
math	414.6 ***	(97.6)	8.767 *	(4.737)
other	-2.9	(63.8)	11.129 ***	(2.499)
Middle-Theoretical				
math	503.6 ***	(111.1)	13.241 **	(5.625)
other	-36.3	(61.2)	11.852 ***	(3.146)
Middle-Practical	74.5	(63.1)	6.143	(5.171)
Vocational	120.1 *	(66.5)	-14.608 ***	(4.436)
Part-time	372.9 ***	(67.8)		

Note: Note: Estimates of a sample of 4927 students or 31932 student-year observations during secondary education. Fixed cost estimates of equation (15). The reported marginal costs of effort are an approximation of the predicted values from the model. All parameters are divided by μ_{time} in equation (15) such that they can be interpreted in minutes of daily travel time. Clas= classical languages included. Math= intensive math. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

A.4 Estimation results long run outcomes

Table A9: Value of obtaining degree

	Degree values	
	coef	**se**
Academic	1481.5 ***	(318.5)
Middle-Theoretical	996.4 ***	(253.5)
Middle-Practical	1173.6 ***	(271.7)
Vocational	594.8 ***	(129.0)
12th grade certificate vocational	852.1 ***	(153.6)

Note: Estimates of μ^{degree} in equation (6). All parameters are divided by μ_{time} in equation (15) such that they can be interpreted in minutes of daily travel time. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A10: Estimation results of long run outcomes (1)

	Higher education					
	Enrollment		First year success		Degree age 25	
	coef	se	coef	se	coef	se
Degree Academic	3.949 ***	(0.261)				
clas+math			-0.015	(0.462)	-0.210	(0.402)
clas			0.433	(0.290)	-0.162	(0.356)
math			0.082	(0.205)	0.313	(0.252)
other			*benchmark*		*benchmark*	
Degree Middle-Theoretical	3.280 ***	(0.237)				
math			-0.669 *	(0.397)	-0.028	(0.471)
other			-0.493 ***	(0.180)	-0.398 *	(0.230)
Degree Middle-Practical	1.960 ***	(0.177)	-0.770 ***	(0.231)	-0.663 **	(0.319)
Degree Vocational	*benchmark*		-1.537 ***	(0.506)	-2.263 ***	(0.539)
Academic level secondary education						
x Academic level higher education	0.539 ***	(0.077)	-0.378 *	(0.209)	-0.266 **	(0.128)
x STEM	-0.788 ***	(0.105)	-0.042	(0.263)	-0.081	(0.288)
Classical languages						
x Academic level higher education	0.807 ***	(0.097)	0.166	(0.192)	0.218	(0.141)
x STEM	-0.684 ***	(0.260)	0.048	(0.396)	0.164	(0.315)
Intensive math						
x Academic level higher education	0.422 ***	(0.093)	-0.024	(0.181)	-0.266 **	(0.128)
x STEM	0.902 ***	(0.164)	0.640	(0.397)	-0.081	(0.288)
Study delay	0.015	(0.180)	-0.119	(0.316)	-1.239 ***	(0.367)
x Academic level higher education	-0.037	(0.100)	0.017	(0.182)	-0.464 **	(0.187)
x STEM	-0.063	(0.128)	-0.238	(0.272)	0.080	(0.239)

Note: Estimates of long run outcomes as specified in section 4.5. Clas= classical languages included. Math= intensive math.

Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A11: Estimation results of long run outcomes (2)

	Higher education						**Unemployment**	
	Enrollment		First year success		Degree age 25		Spell age 25-35	
	coef	se	coef	se	coef	se	coef	se
Male	-0.879 ***	(0.133)	-0.349 *	(0.184)	-0.815 ***	(0.200)	-0.507 ***	(0.076)
x Academic level higher education	0.016	(0.087)	-0.056	(0.149)	0.192	(0.124)		
x STEM	1.725 ***	(0.133)	0.384	(0.290)	1.218 ***	(0.177)		
Language ability	0.130	(0.130)	0.257	(0.173)	0.025	(0.187)	-0.066	(0.057)
x Academic level higher education	0.344 ***	(0.071)	0.269 **	(0.117)	0.334 ***	(0.128)		
x STEM	-0.143	(0.123)	-0.052	(0.308)	-0.148	(0.198)		
Math ability	0.043	(0.102)	0.210	(0.217)	0.111	(0.212)	-0.113 **	(0.051)
x Academic level higher education	0.256 ***	(0.079)	0.407 ***	(0.138)	0.351 **	(0.162)		
x STEM	0.813 ***	(0.162)	-0.077	(0.340)	0.836 ***	(0.254)		
High SES	0.242	(0.158)	0.409 **	(0.161)	0.557 ***	(0.162)	0.080	(0.103)
x Academic level higher education	0.391 ***	(0.057)	0.247 **	(0.101)	0.356 ***	(0.120)		
x STEM	0.220	(0.146)	-0.346	(0.281)	0.026	(0.227)		
Type 2	-0.103	(0.143)	0.482 ***	(0.155)	0.430 **	(0.177)	-0.168	(0.103)
x Academic level higher education	0.337 ***	(0.067)	0.762 ***	(0.138)	1.258 ***	(0.098)		
x STEM	1.693 ***	(0.143)	-0.097	(0.230)	0.739 ***	(0.254)		
Distance (km)	-0.022 ***	(0.002)	-0.003	(0.004)	-0.019 ***	(0.004)		

Note: Estimates of long run outcomes as specified in section 4.5. 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. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A12: Estimation results of long run outcomes (3)

	Higher education		
	Degree age 25		
	coef		se
Same level as enroll	0.154		(0.143)
x Passed enrollment year	2.982	***	(0.164)
Same major as enroll	1.723	***	(0.152)
x Passed enrollment year	0.909	***	(0.198)
Upgraded	-1.576	***	(0.333)
x Passed enrollment year	1.417	***	(0.423)

Note: Estimates of choice-specifc constants long run outcomes as specified in section 4.5. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A13: Estimation results of long run outcomes (4)

	Higher education								
	Enrollment			First year success			Degree age 25		
	coef		se	coef		se	coef		se
Univ Leuven - no STEM	-6.980	***	(0.504)	-1.497		(1.267)	-5.006	***	(0.913)
Univ Leuven - STEM	-9.008	***	(0.495)	-1.946		(1.240)	-6.890	***	(0.992)
Univ Antwerp - No STEM	-9.240	***	(0.548)	-0.792		(1.264)	-7.202	***	(0.989)
Univ Antwerp - STEM	-10.751	***	(0.766)	-0.919		(1.628)	-8.994	***	(1.421)
Univ Brussels - No STEM	-8.532	***	(0.501)	-0.015		(1.136)	-6.472	***	(0.989)
Univ Brussels - STEM	-10.408	***	(0.613)	-1.156		(1.590)	-8.386	***	(1.080)
Univ Ghent - No STEM	-7.591	***	(0.532)	-0.816		(1.194)	-5.502	***	(0.994)
Univ Ghent - STEM	-9.118	***	(0.564)	-2.616	**	(1.334)	-7.136	***	(1.041)
Univ Hasselt - No STEM	-10.663	***	(0.629)	-0.041		(1.554)	-7.159	***	(1.011)
Univ Hasselt - STEM	-9.041	***	(0.457)	-0.881		(1.281)	-8.455	***	(0.982)
Acad college - No STEM	-5.349	***	(0.308)	-0.103		(0.585)	-4.418	***	(0.536)
Acad college - STEM	-5.931	***	(0.295)	-1.165		(0.814)	-4.824	***	(0.669)
Prof college - No STEM	-1.636	***	(0.164)	-0.077		(0.340)	-1.504	***	(0.291)
Prof college - STEM	-3.381	***	(0.212)	0.075		(0.578)	-2.774	***	(0.479)

Note: Estimates of long run outcomes as specified in section 4.5. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A14: Estimation results of long run outcomes (5)

	Unemployment		
	Spell age 25-35		
	coef		se
Secondary education outcomes			
Degree Academic			
clas+math	-0.562		(0.582)
clas	-0.147		(0.565)
math	-0.240		(0.568)
other	-0.201		(0.550)
Degree Middle-Theoretical			
math	-0.794		(0.581)
other	-0.591	*	(0.339)
Degree Middle-Practical	-1.012	***	(0.195)
Degree Vocational	-0.472	***	(0.140)
12th grade certificate vocational	-0.402	**	(0.187)
Part-time track	0.162		(0.155)
Level SE	-0.237		(0.194)
Study delay	0.142	*	(0.075)
Higher education outcomes			
Degree univ Leuven - no STEM	-0.700	***	(0.260)
Degree univ Leuven - STEM	-0.826		(0.638)
Degree univ Antwerp - No STEM	-0.411		(2.799)
Degree univ Antwerp - STEM	0.157		(3.448)
Degree univ Brussels - No STEM	-0.508		(0.810)
Degree univ Brussels - STEM	-0.804		(3.571)
Degree univ Ghent - No STEM	-1.043	**	(0.525)
Degree univ Ghent - STEM	-0.995		(1.385)
Degree univ Hasselt - No STEM	-0.898		(2.809)
Degree univ Hasselt - STEM	-0.254		(2.102)
Degree acad college - No STEM	-0.203		(0.269)
Degree acad college - STEM	-0.617		(0.422)
Degree prof college - No STEM	-0.826	***	(0.125)
Degree prof college - STEM	-0.518	*	(0.289)

Note: Estimates of long run outcomes as specified in section 4.5. Clas= classical languages included. Math= intensive math. Level SE = academic level of study program in secondary education. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A15: Estimation results of long run outcomes (6)

	Higher education			**Unemployment**		
	First year success			Spell age 25-35		
	coef		se	coef		se
Constant	0.033		(0.175)			
Threshold level 2				-0.439	***	(0.167)
Threshold level 3	2.387	***	(0.126)	-0.175		(0.168)
Threshold level 4	4.657	***	(0.397)	0.050		(0.167)
Threshold level 5				0.263		(0.163)
Threshold level 6				0.477	***	(0.162)
Threshold level 7				0.733	***	(0.153)
Threshold level 8				1.024	***	(0.149)
Threshold level 9				1.495	***	(0.158)
Threshold level 10				2.080	***	(0.165)
Academic college						
x Threshold level 3	-0.270		(0.261)			
x Threshold level 4	-0.223		(0.784)			
University						
x Threshold level 3	0.254		(0.227)			
x Threshold level 4	0.245		(0.455)			
STEM						
x Threshold level 3	-0.272		(0.215)			
x Threshold level 4	-0.970	*	(0.526)			

Note: Estimates of long run outcomes as specified in section 4.5. Clas= classical languages included. Math= intensive math. Level SE = academic level of study program in secondary education. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

A.5 Average Treatment Effects on the Treated

Using the parameters of long run outcomes, I can calculate the Average Treatment effects on the Treated (ATT) of the study program a student graduated in as follows:

$$ATT^{j'} = E_{x,\nu,\omega}\left[\tilde{\Psi}_j^w(x_{it_w}(j'),\nu_i,\omega_i^w) - \tilde{\Psi}_j^w(x_{it_w}(j^0),\nu_i,\omega_i^w)|d_{iT_i^{SE}}^{j'} = 1\right] \tag{23}$$

with $E_{x,\nu,\omega}$ an expectations operator over the empirical distribution of the observables x and the estimated distribution of the unobserved types ν and shocks ω. $x_{it_w}(j')$ is the observed state vector of student i in the data and $x_{it_w}(j^0)$ is the same vector but with the graduation track replaced by an arbitrary benchmark program j^0. The ATT then calculates the average effect on w from graduating from j' instead of j^0 for the group of students who graduated from j' in the data. The estimate is a "ceteris paribus" causal effect, i.e. it is the effect of one variable if all other variables that were realized before leaving secondary education are kept fixed. Similarly, I calculate the effect of one year of study delay by comparing outcomes for retained students in the counterfactual scenario where they would not have accumulated study delay. The tables in this section summarize all ATT results.

Table A16: ATTs of high school program and delay in higher education (in %points difference)

		Higher education								
		Enrollment (% of students)			First year success (% of enrolled)			Degree age 25 (% of students)		
Study program		coef		se	coef		se	coef		se
Academic										
	clas+math	+3.31	***	(0.61)	+9.40		(6.70)	+1.61		(2.90)
	clas	+2.69	***	(0.56)	+12.76	***	(4.68)	+3.38		(3.39)
	math	+2.92	***	(0.51)	+6.66		(4.27)	+2.70		(2.59)
	other	*benchmark*			*benchmark*			*benchmark*		
Middle-Theoretical										
	math	+1.38		(1.86)	-1.96		(7.05)	-2.14		(6.12)
	other	-7.81	***	(2.74)	-10.37	**	(4.43)	-11.62	***	(3.56)
Middle-Practical		-31.53	***	(2.92)	-13.77	***	(4.13)	-25.47	***	(3.55)
Vocational		-67.32	***	(2.98)	-25.65	***	(6.34)	-44.21	***	(4.36)
One year of study delay		-0.20		(2.24)	-3.85		(6.59)	-10.37	***	(2.98)
Data		55.2			45.6			39.0		

Note: Average treatment effects on the treated (ATT). Clas= classical languages included. Math= intensive math. Effects on enrollment, first year success and degree completion after graduating from different high school programs, compared to graduating from the academic track without clas or math option, and the effects of one year of study delay, compared to 0. Effects are calculated using indexes, specified in section 4.5, for each individual at the realization of other variables. Effects on obtaining higher education degree at age 25 and on employment spell are total effects, taking into account effects through enrollment and first year performance. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A17: ATTs of high school program and delay on HE degree at age 25: level (in %points difference)

	Academic level higher education					
	University (% of students)		Academic college (% of students)		Professional college (% of students)	
Study program	coef	se	coef	se	coef	se
Academic						
clas+math	+23.81 ***	(4.86)	-3.16 **	(1.26)	-19.04 ***	(3.98)
clas	+19.21 ***	(2.95)	+1.24	(0.78)	-17.06 ***	(3.19)
math	+4.24 *	(2.53)	+2.56 ***	(0.83)	-4.11	(2.85)
other	*benchmark*		*benchmark*		*benchmark*	
Middle-Theoretical						
math	-3.30	(2.22)	+1.40	(1.40)	-0.24	(5.81)
other	-1.43 ***	(0.40)	-1.23 ***	(0.32)	-8.96 ***	(3.35)
Middle-Practical	-1.10 ***	(0.28)	-1.47 ***	(0.28)	-22.91 ***	(3.41)
Vocational	-0.36 ***	(0.12)	-0.88 ***	(0.16)	-42.97 ***	(4.28)
One year of study delay	-0.81 ***	(0.29)	-1.12 ***	(0.30)	-8.43 ***	(2.92)
Data	9.6		4.8		24.8	

Note: Average treatment effects on the treated (ATT). Clas= classical languages included. Math= intensive math. Effects on enrollment, first year success and degree completion after graduating from different high school programs, compared to graduating from the academic track without clas or math option, and the effects of one year of study delay, compared to 0. Effects are calculated using indexes, specified in section 4.5, for each individual at the realization of other variables. Effects on obtaining higher education degree at age 25 and on employment spell are total effects, taking into account effects through enrollment and first year performance. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A18: ATTs of high school program and delay on HE degree at age 25: majors (in %points difference)

	Major					
	STEM (% of students)			No STEM (% of students)		
Study program	coef		se	coef		se
Academic						
clas+math	+5.20		(5.42)	-3.59		(5.66)
clas	-2.16		(2.06)	+5.54	*	(3.32)
math	+11.52	***	(3.12)	-8.82	***	(3.07)
other	*benchmark*			*benchmark*		
Middle-Theoretical						
math	+19.14	***	(4.98)	-21.28	***	(4.41)
other	+2.22	**	(0.99)	-13.85	***	(2.88)
Middle-Practical	+1.81	*	(1.10)	-27.28	***	(3.10)
Vocational	-1.19	*	(0.68)	-43.02	***	(4.11)
One year of study delay	-4.93	***	(1.87)	-5.44	**	(2.34)
Data	10.2			28.8		

Note: Average treatment effects on the treated (ATT). Clas= classical languages included. Math= intensive math. Effects on enrollment, first year success and degree completion after graduating from different high school programs, compared to graduating from the academic track without clas or math option, and the effects of one year of study delay, compared to 0. Effects are calculated using indexes, specified in section 4.5, for each individual at the realization of other variables. Effects on obtaining higher education degree at age 25 and on employment spell are total effects, taking into account effects through enrollment and first year performance. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A19: ATTs of obtaining high school degree and of study delay on unemployment spells (in years)

	Unemployment	
	Spell age 25-35 (mean in years)	
Study program	coef	se
Academic		
clas+math	-0.93	(0.68)
clas	-0.72	(0.72)
math	-0.73	(0.71)
other	-0.76	(0.76)
Middle-Theoretical		
math	-0.85	(0.73)
other	-0.77	(0.72)
Middle-Practical	-0.93	(0.74)
Vocational	-0.37	(0.78)
One year of study delay	+0.30 ***	(0.06)
Data	2.08	

Note: Average treatment effects on the treated (ATT). Clas= classical languages included. Math= intensive math. Effects on enrollment, first year success and degree completion after graduating from different high school programs, compared to finishing high school without a degree (in the academic track without clas or math option), and the effects of one year of study delay, compared to 0. Effects are calculated using indexes, specified in section 4.5, for each individual at the realization of other variables. Effects on obtaining higher education degree at age 25 and on employment spell are total effects, taking into account effects through enrollment and first year performance. Bootstrap standard errors between parentheses. *** p<0.01, ** p<0.05, * p<0.1 (normal-based).

A.6 Counterfactual simulations of removing study programs

In section 7, I discussed the added value estimates of each track and elective course. To derive these estimates, I predicted choices and outcomes from a model without each track and without each elective course and compared them to the status quo. In this section I show the results of these counterfactuals, as well as the biases that arise when study effort is not modeled as a choice variable. Note that the signs of the predicted effects of removing a track or elective course are the opposite of the signs of added value estimates.

Table A20: Predictions of the model: outcomes if tracks are removed

	Remove track			
	Academic	**Middle-theoretical**	**Middle-practical**	**Vocational**
High school graduation (% of students: change in %points)				
Academic	-43.26 *** (2.97)	+8.46 *** (1.43)	+2.62 *** (0.52)	+0.72 *** (0.26)
clas+math	-4.71 *** (1.02)	+0.29 *** (0.10)	+0.11 *** (0.02)	+0.00 (0.01)
clas	-5.05 *** (1.06)	+1.27 *** (0.23)	+0.51 *** (0.11)	+0.12 *** (0.03)
math	-15.60 *** (1.46)	+1.60 *** (0.55)	+0.10 (0.11)	+0.00 (0.05)
other	-17.90 *** (1.78)	+5.30 *** (0.85)	+1.90 *** (0.39)	+0.60 ** (0.24)
Middle-Theoretical	+30.06 *** (2.99)	-11.14 *** (2.12)	+3.96 *** (0.92)	+1.34 *** (0.38)
math	+10.87 *** (1.66)	-2.13 *** (0.66)	+0.27 ** (0.11)	+0.05 (0.04)
other	+19.19 *** (2.69)	-9.01 *** (1.56)	+3.69 *** (0.85)	+1.29 *** (0.37)
Middle-Practical	+8.20 *** (1.74)	+1.40 ** (0.65)	-11.20 *** (2.05)	+8.40 *** (0.80)
Vocational	+4.26 *** (0.73)	+2.00 *** (0.27)	+5.06 *** (0.59)	-21.30 *** (0.94)
Dropout	+0.68 *** (0.25)	-0.91 *** (0.21)	-0.46 (0.37)	+10.72 *** (1.03)
Students with grade retention	+3.10 ** (1.45)	-2.70 *** (0.39)	-1.90 *** (0.38)	+9.40 *** (0.68)
Higher education (% of students: change in %points)				
Enrollment	-8.53 *** (1.19)	-0.28 (0.43)	+0.27 (0.33)	+3.72 *** (0.46)
First year successful (among enrolled)	-3.21 (2.83)	+0.73 (0.52)	+0.42 (0.36)	-0.75 *** (0.29)
Degree (age 25)	-8.64 *** (1.37)	+0.62 (0.39)	+0.57 ** (0.28)	+1.90 *** (0.30)
University degree	-4.68 *** (0.62)	+0.34 *** (0.09)	+0.10 *** (0.03)	+0.02 ** (0.01)
Academic college degree	-1.21 *** (0.25)	+0.16 *** (0.05)	+0.07 ** (0.03)	+0.03 ** (0.01)
Professional college degree	-2.75 ** (1.14)	+0.12 (0.36)	+0.40 (0.25)	+1.85 *** (0.29)
Degree in STEM major	+1.31 (0.81)	-0.24 * (0.14)	-0.20 ** (0.09)	+0.15 ** (0.07)
Unemployment (mean in years: difference in means)				
Spell age 25-35	+0.02 (0.05)	-0.01 (0.01)	+0.03 * (0.02)	+0.09 ** (0.04)

Note: Predictions from a dynamic model with program and effort choice in secondary education. Differences between outcomes in a world where tracks are removed with the status quo. Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A21: Predictions of the model: outcomes if elective courses are removed

		Remove elective course				
		Classical languages			Intensive math	
High school graduation (% of students: change in %points)						
Academic		-4.36	***	(0.72)	-1.06		(0.74)
	clas+math	-4.71	***	(1.02)	-4.71 ***	(1.02)
	clas	-5.05	***	(1.06)	8.55 ***	(1.35)
	math	4.70	***	(1.14)	-15.60 ***	(1.46)
	other	0.70		(0.72)	10.70 ***	(1.42)
Middle-Theoretical		2.07	***	(0.52)	-0.84 *	(0.47)
	math	0.48	**	(0.19)	-2.13 ***	(0.66)
	other	1.59	***	(0.39)	1.29	(0.81)
Middle-Practical		0.90	***	(0.25)	1.50 ***	(0.54)
Vocational		1.03	***	(0.18)	0.34 **	(0.17)
Dropout		0.31	***	(0.11)	-0.09	(0.11)
Students with grade retention		2.90	***	(0.62)	-2.10 ***	(0.81)
Higher education (% of students: change in %points)						
Enrollment		-1.82	***	(0.24)	-1.14 ***	(0.25)
First year successful (among enrolled)		-0.60		(0.58)	1.39	(1.41)
Degree (age 25)		-1.85	***	(0.38)	-0.18	(0.66)
University degree		-1.92	***	(0.44)	0.39	(0.65)
Academic college degree		-0.16	*	(0.09)	-0.47 ***	(0.17)
Professional college degree		0.23		(0.36)	-0.10	(0.67)
Degree in STEM major		0.40		(0.24)	-2.49 ***	(0.75)
Unemployment (mean in years: difference in means)						
Spell age 25-35 (mean in years)		+0.03	**	(0.01)	+0.01	(0.03)

Note: Predictions from a dynamic model with program and effort choice in secondary education. Differences between outcomes in a world where tracks are removed with the status quo. Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A22: Bias in predictions of the model: tracks removed (1)

	Remove track					
	Academic			**Middle-Theoretical**		
	With effort	Bias without effort		With effort	Bias without effort	
High school graduation (% of students: change in %points)						
Academic	-43.26	2.07 **	(0.91)	8.5	0.66	(0.45)
clas+math	-4.71	-0.18	(0.24)	0.3	0.13 *	(0.08)
clas	-5.05	-0.15	(0.39)	1.3	0.03	(0.14)
math	-15.60	0.50	(0.58)	1.6	0.30	(0.21)
other	-17.90	1.90 ***	(0.67)	5.3	0.20	(0.33)
Middle-Theoretical	+30.06	-0.99	(0.88)	-11.1	-1.19 *	(0.61)
math	+10.87	0.53	(0.45)	-2.1	-0.37 *	(0.22)
other	+19.19	-1.52 **	(0.71)	-9.0	-0.82 *	(0.47)
Middle-Practical	+8.20	-1.30 *	(0.70)	1.4	0.70 **	(0.28)
Vocational	+4.26	0.07	(0.24)	2.0	-0.08	(0.19)
Dropout	+0.68	0.11	(0.20)	-0.9	0.15	(0.15)
Students with grade retention	+3.10	-2.50 ***	(0.65)	-2.7	1.20 ***	(0.31)
Higher education outcomes (% of students: change in %points)						
Enrollment	-8.53	0.40	(0.30)	-0.3	-0.25 *	(0.15)
First year successful (among enrolled)	-3.21	0.47 *	(0.24)	0.7	-0.04	(0.14)
Degree (age 25)	-8.64	0.89 ***	(0.30)	0.6	-0.40 ***	(0.11)
University degree	-4.68	0.10	(0.15)	0.3	0.02	(0.03)
Academic college degree	-1.21	0.23 ***	(0.06)	0.2	-0.02	(0.02)
Professional college degree	-2.75	0.56 **	(0.24)	0.1	-0.40 ***	(0.11)
Degree in STEM major	+1.31	0.26 *	(0.16)	-0.2	-0.12 **	(0.05)
Unemployment (mean in years: difference in means)						
Spell age 25-35	+0.02	-0.01	(0.01)	-0.01	0.01 **	(0.01)

Note: Predictions from a dynamic model with program and effort choice in secondary education and the bias in a model that does not allow for effort to be a choice. Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A23: Bias in predictions of the model: tracks removed (2)

	Remove track					
	Middle-Practical			Vocational		
	With effort	Bias without effort		With effort	Bias without effort	
High school graduation (% of students: change in %points)						
Academic	+2.62	-0.30	(0.25)	0.7	0.24 *	(0.14)
clas+math	+0.11	0.04	(0.03)	0.0	0.00	(0.01)
clas	+0.51	-0.04	(0.09)	0.1	-0.06 **	(0.03)
math	+0.10	0.10	(0.09)	0.0	0.10	(0.07)
other	+1.90	-0.40 *	(0.21)	0.6	0.20	(0.13)
Middle-Theoretical	+3.96	0.72 **	(0.31)	1.3	0.33	(0.22)
math	+0.27	0.14	(0.10)	0.1	0.05	(0.04)
other	+3.69	0.58 *	(0.31)	1.3	0.28	(0.22)
Middle-Practical	-11.20	-0.30	(0.47)	8.4	-0.90 **	(0.38)
Vocational	+5.06	-0.24	(0.22)	-21.3	0.06	(0.39)
Dropout	-0.46	0.15	(0.22)	10.7	0.44	(0.31)
Students with grade retention	-1.90	0.80 **	(0.34)	9.4	2.60 ***	(0.35)
Higher education outcomes (% of students: change in %points)						
Enrollment	+0.27	-0.09	(0.17)	3.7	-0.26 *	(0.15)
First year successful (among enrolled)	+0.42	-0.06	(0.08)	-0.7	0.14 **	(0.07)
Degree (age 25)	+0.57	-0.23 *	(0.12)	1.9	-0.25 **	(0.11)
University degree	+0.10	-0.00	(0.02)	0.0	-0.01	(0.01)
Academic college degree	+0.07	-0.01	(0.02)	0.0	0.00	(0.01)
Professional college degree	+0.40	-0.22 **	(0.11)	1.9	-0.24 **	(0.10)
Degree in STEM major	-0.20	-0.02	(0.03)	0.1	-0.03	(0.02)
Unemployment (mean in years: difference in means)						
Spell age 25-35	+0.03	0.01 *	(0.00)	+0.09	0.02 ***	(0.01)

Note: Predictions from a dynamic model with program and effort choice in secondary education and the bias in a model that does not allow for effort to be a choice. Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

Table A24: Bias in predictions of the model: elective courses removed

	Remove elective course						
	Classical languages			Intensive math			
	With effort	Bias without effort		With effort	Bias without effort		
High school graduation (% of students: change in %points)							
Academic	-4.36	0.17	(0.34)	-1.1	-0.43		(0.35)
clas+math	-4.71	-0.18	(0.24)	-4.7	-0.18		(0.24)
clas	-5.05	-0.15	(0.39)	8.6	0.95	*	(0.50)
math	+4.70	0.50	(0.31)	-15.6	0.50		(0.58)
other	+0.70	0.00	(0.34)	10.7	-1.70	**	(0.69)
Middle-Theoretical	+2.07	0.00	(0.24)	-0.8	0.31		(0.25)
math	+0.48	0.22 **	(0.10)	-2.1	-0.37	*	(0.22)
other	+1.59	-0.22	(0.19)	1.3	0.68	**	(0.27)
Middle-Practical	+0.90	-0.10	(0.15)	1.5	0.40	*	(0.21)
Vocational	+1.03	0.07	(0.11)	0.3	0.04		(0.12)
Dropout	+0.31	-0.01	(0.09)	-0.1	-0.17		(0.12)
Students with grade retention	+2.90	-0.20	(0.26)	-2.1	-0.60		(0.71)
Higher education outcomes (% of students: change in %points)							
Enrollment	-1.82	0.03	(0.11)	-1.1	0.02		(0.14)
First year successful (among enrolled)	-0.60	-0.04	(0.08)	1.4	0.17		(0.25)
Degree (age 25)	-1.85	0.10	(0.12)	-0.2	0.05		(0.27)
University degree	-1.92	-0.10	(0.08)	0.4	0.18		(0.24)
Academic college degree	-0.16	0.06	(0.04)	-0.5	0.04		(0.06)
Professional college degree	+0.23	0.14	(0.10)	-0.1	-0.17		(0.16)
Degree in STEM major	+0.40	0.08	(0.08)	-2.5	0.08		(0.18)
Unemployment (mean in years: difference in means)							
Spell age 25-35 (mean in years)	+0.03	0.00	(0.00)	+0.01	-0.01		(0.01)

Note: Predictions from a dynamic model with program and effort choice in secondary education and the bias in a model that does not allow for effort to be a choice. Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).

A.7 Other tables

Table A25: 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	10.1	6.5	4.0	0.9	2.1
grade 9+10	9.6	6.0	4.5	4.3	1.6
grade 11+12	6.7	6.7	6.7	6.7	0
Middle-Theoretical					
grade 7+8	30.6	26.0	19.3	1.0	1.3
grade 9+10	100	17.9	11.6	6.2	1.3
grade 11+12	100	11.1	11.1	11.1	0
Middle-Practical					
grade 7+8	38.5	32.3	31.5	3.7	3.8
grade 9+10	100	100	22.3	9.7	0
grade 11+12	100	100	14.9	14.9	0
Vocational					
grade 7+8	100	100	100	6.1	0
grade 9+10	100	100	100	12.8	0
grade 11+12+13	100	100	100	13.0	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 A26: Transition matrix (in % of students)

	Last choice						
	Acad-clas	Acad-no clas	Middle-theo	Middle-prac	Vocational	Dropout	
First choice							
Acad-clas	9.3	15.5	4.6	2.6	1.6	1.5	35.1
Acad-no clas		10.1	6.3	5.1	3.3	2.4	27.1
Middle-theo		1.3	4.5	3.9	6.7	3.5	19.9
Middle-prac		0.1	1.0	1.1	3.2	1.8	7.3
Vocational					5.6	4.9	10.6
	9.3	27.0	16.4	12.7	20.5	14.2	100.0

Proportion math in Academic-class (last choice): 40.4

Proportion math in Academic-no class (last choice): 46.7

Proportion math in Middle-theoretical (last choice): 15.6

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

Table A27: Predictions of the model: substitution patterns counterfactual tracking policies

	Status quo	Policy change B-certificate			
		Repeat		Downgrade	
High school graduation (% of students and change in %points)					
Academic	*43.3*	+1.00 **	(0.46)	-1.53 ***	(0.49)
clas+math	4.7	+0.01	(0.06)	+0.22 **	(0.10)
clas	5.1	+0.09	(0.20)	+0.25 **	(0.12)
math	15.6	+0.00	(0.13)	-1.20 ***	(0.26)
other	17.9	+0.90 **	(0.38)	-0.80 **	(0.33)
Middle-Theoretical	*11.1*	-1.46 ***	(0.31)	-0.83 *	(0.45)
math	2.1	-0.46 ***	(0.10)	+0.10	(0.15)
other	9.0	-1.00 ***	(0.29)	-0.93 ***	(0.35)
Middle-Practical	*11.2*	-2.85 ***	(0.40)	-0.10	(0.37)
Vocational	*21.3*	-0.88 ***	(0.23)	+3.83 ***	(0.30)
Dropout	*13.2*	+4.03 ***	(0.35)	-1.45 ***	(0.18)

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

Table A28: Bias in predictions of the model: tracking policy

	Policy change B-certificate					
	Repeat			**Downgrade**		
	With effort	Bias without effort		With effort	Bias without effort	
High school graduation (% of students: change in %points)						
Academic	+1.00	-0.73 **	(0.34)	-1.53	-0.67 **	(0.28)
clas+math	+0.01	0.10	(0.07)	+0.22	-0.03	(0.09)
clas	+0.09	0.07	(0.15)	+0.25	-0.14	(0.12)
math	+0.00	-0.20	(0.14)	-1.20	-0.20	(0.17)
other	+0.90	-0.70 ***	(0.25)	-0.80	-0.30	(0.20)
Middle-Theoretical	-1.46	-0.22	(0.26)	-0.83	-0.04	(0.23)
math	-0.46	0.03	(0.10)	+0.10	0.01	(0.11)
other	-1.00	-0.25	(0.23)	-0.93	-0.05	(0.18)
Middle-Practical	-2.85	-0.52 **	(0.22)	-0.10	0.00	(0.23)
Vocational	-0.88	0.01	(0.26)	+3.83	0.62 ***	(0.20)
Dropout	+4.03	1.55 ***	(0.27)	-1.45	0.13	(0.15)
Students with grade retention	+9.50	3.10 ***	(0.51)	-9.60	1.00 **	(0.42)
Higher education (% of students: change in %points)						
Enrollment	-1.90	-1.16 ***	(0.20)	-1.21	-0.51 ***	(0.11)
First year successful (among enrolled)	+0.11	0.08	(0.12)	+0.81	-0.17	(0.12)
Degree (age 25)	-1.75	-0.94 ***	(0.19)	+0.32	-0.66 ***	(0.09)
University degree	-0.13	-0.05	(0.05)	+0.30	-0.14 ***	(0.05)
Academic college degree	-0.16	-0.06 **	(0.03)	+0.09	-0.07 ***	(0.02)
Professional college degree	-1.46	-0.83 ***	(0.14)	-0.07	-0.45 ***	(0.08)
Degree in STEM major	-0.48	-0.16 ***	(0.05)	+0.27	-0.10 **	(0.05)
Unemployment (mean in years: difference in means)						
Spell age 25-35	+0.12	0.05 ***	(0.01)	-0.05	0.01 **	(0.00)

Note: Predictions from a dynamic model with program and effort choice in secondary education and the bias in a model that does not allow for effort to be a choice. B-certificate = students acquired skills to proceed to next grade but only in track of lower academic level or if they drop elective course. Status quo = students can choose to downgrade or repeat grade after obtaining B-certificate, Repeat = students must repeat grade after obtaining B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining B-certificate. Bootstrap standard errors between parentheses. p<0.01, ** p<0.05, * p<0.1 (normal-based).