WORKING PAPERS

N° 1002

June 2023

"Dynamic Effort Choice in High School: Costs and Benefits of an Academic Track"

Olivier De Groote

Toulouse School of Economics

Dynamic Effort Choice in High School: Costs and Benefits of an Academic Track

Olivier De Groote

June 27, 2023

Abstract

I investigate high school tracking policies using a dynamic discrete choice model of study programs and unobserved effort. I estimate the model using data from Flanders (Belgium) and perform an ex-ante evaluation of a policy that encourages underperforming students to switch to less academically oriented programs. This reduces grade retention by a third and dropout by 11%. Although it decreases college enrollment, the decrease in college graduation is small and insignificant. I also show that modeling effort is important, otherwise we would predict smaller decreases in grade retention and dropout and larger decreases in college enrollment and graduation.

Olivier De Groote: Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France. E-mail: olivier.de-groote@tse-fr.eu. This paper benefited from helpful comments at various stages from Peter Arcidiacono, Estelle Cantillon, Jan De Loecker, Koen Declercq, Eric French, Vishal Kamat, Thierry Magnac, Arnaud Maurel, François Poinas, Uta Schönberg, Jo Van Biesebroeck, Frank Verboven and several audiences at KU Leuven and Duke University. It also benefited from discussions during and after presentations at ULiège, RUG Groningen, UAB Barcelona, Toulouse School of Economics, University of Cambridge, LMU Munich, McGill University, UNC Chapel Hill, Oslo University, ENSAE-CREST, Toronto Rotman School of Management, University of Alicante, Helsinki GSE, IWAEE 2017, EALE 2017, ECORES and LEER summer schools 2017, NESG 2018, CEPR Applied IO 2018, ES 2018 European Winter Meeting, SOLE 2019, Barcelona GSE summer forum 2019, IZA/SOLE Transatlantic Meeting for Labor Economists 2020 and the EEA 2020. I also thank Jan Van Damme, Bieke De Fraine, and the Flemish Ministry of Education and Training, for providing the LOSO dataset. Funding for this project was generously provided by the Research Foundation Flanders (FWO). Olivier De Groote acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

1 Introduction

Secondary education either prepares students for higher education (academic curricula) or the job market (vocational curricula). The European 2020 target of 40% college-educated people induces many countries to place students in an academic track. In the US, there is a similar trend toward more academic course taking, especially in STEM (Science, Technology, Engineering, Math)-fields (Nord et al., 2011). While such courses could help to prepare for college (Guyon et al., 2012; Joensen and Nielsen, 2009), it is unclear whether it is desirable to promote such courses to a large group of students. Many students could experience difficulties in completing them successfully, thereby risking low grades, grade retention and dropout.

I study the heterogeneous impact of an academic track in Flanders, the largest region of Belgium. While many countries offer a variety of courses to cater to students with diverse interests and abilities, this context takes it a step further by institutionalizing the process through a transparent tracking system. This facilitates a straightforward examination of the trade-offs that students encounter. At the age of 12, students enter high school and choose a study program: a bundle of one of four tracks and a few elective courses. Tracking occurs early, but gradually, as many students later switch to a program of lower academic level, also known as "downgrading". While it is often a choice, it can also be imposed on them. At the end of the year, students obtain a certificate based on their performance which defines their choice set the year after. The best outcome is an A-certificate, keeping all options open. Bad performance leads to a C-certificate, requiring them to repeat the grade. Of particular interest is the intermediate performance outcome: a B-certificate. These students did not perform well on important courses of the program, but they are allowed to transition to the next grade if they downgrade. To stay in the same program, they have to repeat the grade. 35% of students receive this certificate at some point during secondary education.

To analyze this trade-off, I first estimate a model of students' decisions and outcomes during high school and higher education. I then use the estimates to simulate changes to the B-certificate policy. First, I show the importance of allowing students to downgrade by simulating a context in which underperformance inevitably leads to grade retention. Next, I perform an ex-ante analysis of a new policy in Flanders by no longer allowing students to repeat the grade if they could downgrade instead. The results are in favor of such a policy, showing that the total costs of grade repetition outweigh the benefits of a better track.

The main part of the model rationalizes two yearly decisions of forward-looking high school students. First, they make a discrete choice between study programs. Second, they make a continuous choice on the level of effort to exert. Since I do not observe measures of study effort, I allow students to choose *effective* study effort in the form of the odds of avoiding a bad performance outcome. This effort variable determines end-of-year performance up to an unanticipated shock. Flow utility depends on a fixed cost and a variable cost of effort, with constant marginal costs. To identify marginal costs, I exploit an Euler equation for effort that arises naturally in dynamic models. Fixed costs then rationalize the program decisions. Given fixed and marginal costs, students can re-optimize both program choices and effort levels in counterfactual simulations. Utilities and state transitions of a model without effort choice (i.e. a pure discrete choice model as in Rust (1987); Magnac and Thesmar (2002)) serve as inputs for the identification of marginal and fixed costs. Therefore, it is straightforward to apply standard approaches to identify a model with persistent unobserved heterogeneity (Hu and Shum, 2012; Kasahara and Shimotsu, 2009) and estimate it without solving it using Conditional Choice Probability (CCP) estimation and finite dependence (Arcidiacono and Miller, 2011; Hotz and Miller, 1993).

The model also includes higher education enrollment and graduation such that I can estimate how they are influenced by high school study programs and grade retention. I control for gender, socio-economic status (SES), two continuous measures of cognitive ability at the start of high school and two unobserved types. These controls also affect fixed and marginal costs during high school, thereby allowing for general forms of non-random selection in programs and grade retention. Unobserved heterogeneity can be identified nonparametrically by exploiting rich panel data (Hu and Shum, 2012) and by using travel time to high school programs as an exclusion restriction (Heckman and Navarro, 2007).

The estimates reveal the channels that will drive the counterfactual results and are important to consider when designing an optimal policy. First, policies should not ignore how much has been determined before. Students' initial conditions at age 12 strongly influence the costs and benefits during high school. For example, a 10% of a standard deviation increase in language ability lowers the marginal cost of effort in the vocational track by 6% and in the academic track by 9%. It also decreases the fixed costs of attending a more acadamically oriented program and increases the expected utility from attending college. We see qualitatively similar effects for math ability, female and high SES, as well as strong differences between unobserved types. Second, there are large fixed costs of both repeating grades and switching tracks. While marginal costs decrease in academically oriented programs during a repeated year, accumulated study delay leads to higher costs later on. Finally, graduating from an academically oriented program increases the likelihood of obtaining a higher education degree while repeating a grade decreases this.

The counterfactual simulations show that the benefits of an academically oriented program are insufficient to make up for the harm of grade retention. The policy that forces students to "repeat" the grade if they obtain a B-certificate increases the number of retained students, causing a one-third increase in high school dropouts, and a 4% decrease in the number of college graduates. The ex-ante evaluation of the policy that would force these students to "downgrade" instead decreases retained students by a third and dropouts by 11%. While it also decreases higher education enrollment by 2.4%, the impact on the number of college graduates is smaller than 1% and not statistically significant. Note that both policies incentivize students to exert more effort to avoid bad performance outcomes, particularly those at risk. One of the results is that the number of students directly affected by the policy change decreases. The share of students who obtained at least one B-certificate decreases by 10.2 % points in the "repeat" policy and by 3.5 % points in the "downgrade" policy. I compare my findings to those of a pure discrete choice model in the spirit of Rust (1987). In this model, students still choose a study program but observed performance is modeled as a policy-invariant function of observable characteristics and unobserved types, i.e. this ignores counterfactual changes in effort. Here we see much smaller decreases in the number of B-certificates (respectively 6.3 % points and 0.9 % points). Moreover, there are larger increases in grade retention and drop out in the "repeat" policy and smaller decreases in the "downgrade" policy. Importantly, without effort in the model, we do predict a statistically significant decrease in the number of college graduates in the "downgrade" policy of 2.5%.

A welfare analysis shows that the "downgrade" policy leads to a small loss for students. This can be explained by the reduction of their choice set. However, this loss is largely offset by the taxpayers' gains through reduced spending on education and increased tax returns. The big impact of initial conditions on effort costs suggests investing these gains in early childhood education. Students with higher ability at the start of secondary education are more able to invest in an academically oriented program. These dynamic complementarities avoid the trade-off that arises from the status quo remedial strategy during high school where underperforming students need to repeat a grade to continue in the academic program. While the effects of the downgrade policy are not very different across groups, such a shift of resources is also expected to reduce the influence of SES on long-run outcomes. A decomposition exercise shows that nearly half of the SES gap in college can be mitigated by bringing math and language ability at age 12 to the same level. Additional gains are possible if it also improves unobserved measures of ability.

Related literature This paper contributes to three strands of literature.

First, it contributes to the literature on the returns to educational investments. Altonji et al. (2012) review the literature on the effects of high school curriculum on educational attainment and wages, initiated by Altonji (1995). The literature has found positive effects of intensive math courses (Aughinbaugh, 2012; Goodman, 2019; Joensen and Nielsen, 2009; Rose and Betts, 2004), and stressed the importance of comparative advantages when comparing returns to academic and vocational curricula (Kreisman and Stange, 2020; Meer, 2007). Selection into a beneficial program is not random, explaining why investing in early childhood education is effective because it induces students to opt for a better program later through dynamic complementarities (Cunha and Heckman, 2009; Cunha et al., 2010; Heckman and Mosso, 2014). A separate literature looks at the causal impact for a student of being retained in school (Cockx et al., 2019; Fruehwirth et al., 2016; Gary-Bobo et al., 2016; Jacob and Lefgren, 2009; Manacorda, 2012). I contribute to this literature by jointly analyzing high school program choice and grade retention within a structural model. This approach has several benefits: (1) I can simulate a policy that goes beyond interpreting treatment effects, as I can take into account how the threat of grade retention affects students' effort decisions, (2) dynamic complementarities are illustrated through counterfactual policies, by comparing remedial strategies in high school to the effects of initial conditions, (3) I identify new parameters that quantify the cost of grade retention and the differences in effort costs between students with different initial conditions.

Second, it contributes to the literature on educational tracking policies. Cummins (2017); Duflo et al. (2011); Fu and Mehta (2018); Hanushek and Woessmann (2006); Pekkarinen et al. (2009) and Roller and Steinberg (2020) look at the impact of tracking students at an early age. Dustmann et al. (2017) and Guyon et al. (2012) investigate the long-run impact of the academic track for specific groups. Cockx et al. (2019) estimate average treatment effects within high school for students that are forced to repeat grades or switch tracks. Recent evidence suggests that switching tracks can reduce the negative impact of early track choice (De Groote and Declercq, 2021; Dustmann et al., 2017). I contribute by investigating the impact of flexibility in tracking policies during secondary education.

Finally, this paper contributes to the estimation of dynamic discrete choice models of educational decision-making. Dynamic models in the spirit of Rust (1987) typically involve discrete choices and stochastic state transitions, where the choice-specific utility functions and the distribution of state transitions are considered to be policy-invariant. Since Keane and Wolpin (1997), dynamic discrete choice models have often been used to model student behavior, many of which include performance measures as an exogeneous stochastic process: course grades (Arcidiacono, 2004; Arcidiacono et al., 2023; Eckstein and Wolpin, 1999), course credits (Declercq and Verboven, 2018; Joensen and Mattana, 2021), college admission probabilities (Arcidiacono, 2005) or length of study (Beffy et al., 2012). Without access to effort data, a policy-invariant distribution of performance excludes changes in effort in counterfactual simulations. This is inconsistent with evidence from theory (Costrell, 1994), field experiments (Dubois et al., 2012), and natural experiments (Garibaldi et al., 2012). Others have used observable measures of study effort in the model (Ahn et al., 2022; Fu and Mehta, 2018; Todd and Wolpin, 2018). I show that data on program choices and performance outcomes are sufficient to identify policy-invariant cost parameters that allow for effort changes in counterfactuals. As in Hu and Xin (2022), I include an unobserved choice variable that influences state transitions. However, I do not impose an exclusion restriction in the state transition rule but make use of a first-order condition coming from the dynamics in the model. This is similar to modeling optimal job-finding rates (Cockx et al., 2018; Paserman, 2008; van den Berg and van der Klaauw, 2019) or unobserved consumption-savings decisions (Gayle et al., 2015; Gayle and Miller, 2015; Margiotta and Miller, 2000). I apply this idea in the dynamic discrete choice framework of Rust (1987), Hotz and Miller (1993) and Arcidiacono and Miller (2011):

I formulate general identification and estimation strategies, including extensions to allow for multiple unobserved choice variables and effects beyond the next period state variable.

The rest of the paper is structured as follows. Section 2 describes the methodological contribution of the paper using a two-period binary choice model. Section 3 describes the institutional context, the data, and policy issues, and section 4 applies the model to the data. I discuss the estimation results in section 5 and I simulate tracking policies in section 6. Finally, I conclude in section 7.

2 The cost of effort

I describe a model of discrete study program decisions and continuous effort decisions and how it can be estimated using only data on program choices and performance outcomes. I discuss the simplest case by letting students choose whether they want to continue in school or drop out, both in period 1 (high school) and period 2 (college). Performance enters as a dummy equal to one when a student obtains a high school degree, which is required to access college.

The main text demonstrates the identification of fixed and marginal costs of effort using the CCPs of program decisions and the observed state transitions of performance outcomes. In Appendix A, a more comprehensive proof is presented, which is summarized at the end of this section. Importantly, if CCPs and state transitions can be identified for different unobserved types (as in Hu and Shum (2012) and Kasahara and Shimotsu (2009)), then fixed and marginal costs in high school, as well as the expected utility of enrolling in college can depend on these types as well. This is essential to generate effort choices that vary by unobserved ability.

2.1 Model

Consider a student i before entering the final year of high school in period $t = 1$. The student can finish high school ($j = 1$) or drop out ($j = 0$). If i stays in school, flow utility is

$$u_i + \varepsilon_{ij1} = u(x_i, y_i) + \varepsilon_{ij1} \tag{1}$$
$$\text{with } u(x_i, y_i) = -C^0(x_i) - c(x_i)y_i$$

x_i is a vector of time-invariant state variables, known to the econometrician and the student (such as observed measures of ability or SES). $y_i \in (0, +\infty)$ is the effective study effort, a choice variable that influences the performance distribution. ε_{ijt} is an extreme value type 1 taste shock, unobserved by the econometrician but observed by the student. In the spirit of Keane and Wolpin (1997), I call $-\left(u(x_i, y_i) + \varepsilon_{ij1}\right)$ the effort cost of schooling. It consists of a fixed cost ($C^0(x_i)$), and a variable cost, rising in the level of effort at marginal cost $c(x_i) > 0$.

After $t = 1$, i can obtain a high school degree: a dummy state variable that is only available in $t = 2$: g_i. I allow for uncertainty and assume students expect to obtain a degree with probability

$$\phi(y_i) = \frac{y_i}{1 + y_i}.$$

This explains why we call y_i *effective* study effort. It is a tool to choose the distribution of performance.[1] y_i can be interpreted as the odds of obtaining a high school degree ($y_i = \frac{\phi(y_i)}{1-\phi(y_i)}$) and $c(x_i)$ as the marginal cost to increase them by one unit. The marginal costs depend on how well students are prepared, how much they

[1] Alternatively, we could write $y_i = \mathcal{Y}_i(e_i)$ with e_i a vector of all measures of effort that matter for performance and $\mathcal{Y}_i(.)$ a production function. Students with high marginal costs experience higher costs of increasing y_i. This could be explained by a worse technology $\mathcal{Y}_i(.)$ (less help from parents, lower ability...) or a stronger distaste to increase e_i (motivation). For counterfactuals choices and welfare we only require y_i, avoiding the need to disentangle both.

(dis)like studying and the constraints they face in performing well. It implies a cost function that is increasing and convex in the probability to obtain a good performance outcome.[2] The fixed cost captures a distaste to go to school, regardless of expected performance (such as travel costs or social norms).

Students who drop out ($j = 0$) in $t = 1$ never return to school and receive a lifetime utility of dropping out. In $t = 2$, students with a high school degree can stay in school ($j = 1$) by going to college, or drop out. The lifetime utility after leaving high school is specified as $\Psi_j(x_i) + \varepsilon_{ij2}$. I do not distinguish between channels of utility in $t = 2$ and consider $\Psi_0(.)$ and $\Psi_1(.)$ to be policy-invariant functions of x_i. To keep this model simple, I make several (strong) assumptions that will be relaxed in the application. I do not allow for grade retention by assuming that students who failed have to drop out, and I assume $\Psi_0(x_i, g_i) \equiv \Psi_0(x_i)$. This implies that a high school degree only has value when it is used to get into college. As only differences in utility are identified, we need to treat $\Psi_0(x_i)$ as known (Magnac and Thesmar, 2002). I set $\Psi_0(x_i) = 0$ and interpret $\Psi_1(x_i)$ as the difference in expected lifetime utility.

2.2 Solution

In $t = 2$, i has a choice only after obtaining a high school degree ($g_i = 1$). Since $t = 2$ is the final period, this is equivalent to a static model. The student goes to college if $\Psi_1(x_i) + \varepsilon_{i12} > \varepsilon_{i02}$. Let d_{it} be the chosen option by i at time t. The probability to go to school for students with a high school degree is:

$$\Pr(d_{i2} = 1 | x_i, g_i = 1) = \frac{\exp(\Psi_1(x_i))}{1 + \exp(\Psi_1(x_i))} \ . \tag{2}$$

In period 1, the problem is dynamic. Students do not know if they will be successful, but they know the distribution of g_i. There is also uncertainty about future

[2]Let $\bar{g}_i = \phi(y_i)$, then $\frac{du(x_i,y_i)}{d\bar{g}_i} = \frac{\partial u_i}{\partial y_i}\frac{d\phi^{-1}(\bar{g}_i)}{d\bar{g}_i} < 0$ with $\frac{\partial u_i}{\partial y_i} = -c(x_i) < 0$ and $\frac{d\phi^{-1}(\bar{g}_i)}{d\bar{g}_i} = \frac{1}{(1-\bar{g}_i)^2} > 0$, $\frac{d^2 u(x_i,y_i)}{d\bar{g}_i^2} = -c(x_i)\frac{2}{(1-\bar{g}_i)^3} < 0$.

taste shocks. The lifetime utility of drop out is given by $\Psi_0(x_i) + \varepsilon_{i01} = \varepsilon_{i01}$. The lifetime utility of choosing the high school option is represented by the conditional value function (added with taste shock ε_{i11}):

$$v(x_i, y_i) = u(x_i, y_i) + \beta\gamma + \beta\phi(y_i) \ln(1 + \exp\Psi_1(x_i, g_i = 1)) \tag{3}$$

with γ the Euler constant, $\beta \in (0, 1)$ the one-period discount factor and $\ln(1 + \exp\Psi_1(x_i, g_i = 1))$ the logsum expression, net of γ.

To find the optimal y_i, we maximize v, giving the following first-order condition (FOC):

$$\frac{dv(x_i, y_i)}{dy_i} = \frac{\partial u(x_i, y_i)}{\partial y_i} + \beta\left(\frac{\partial\phi(y_i)}{\partial y_i} \ln(1 + \exp\Psi_1(x_i))\right) = 0 \text{ if } y_i = y_i^* \tag{4}$$

with y_i^* the optimal choice of y_i, $\frac{\partial u(x_i,y_i)}{\partial y_i} = -c(x_i)$ and $\frac{\partial\phi(y_i)}{\partial y_i} = (1 + y_i)^{-2}$. This FOC is an Euler equation that equalizes today's marginal cost of effort to its discounted marginal benefit in the next period such that

$$y_i^* = \sqrt{\frac{\beta\ln(1 + \exp\Psi_1(x_i))}{c(x_i)}} - 1. \tag{5}$$

The optimal level of effort increases in the discounted surplus of being able to enter college ($\beta\ln(1 + \exp\Psi_1(x_i))$) and decreases in its marginal cost $c(x_i)$. Requiring an interior solution for y_i puts an upper bound on the latter.

2.3 Identification of fixed and marginal costs

I first discuss the comparison of this model to a "pure" dynamic discrete choice model and show that its utilities and state transitions are equivalent to the equilibrium outcomes in the proposed model on effort. I then discuss how they can be used to identify fixed and marginal costs. Finally, I discuss intuition in a more general setting and refer to Appendix A for the proof.

2.3.1 The pure discrete choice model

Define a "pure" dynamic discrete choice model in line with Rust (1987). I.e. there is a discrete choice (schooling), and a stochastic state (performance) that evolves exogenously as a function of x_i which can be estimated from the data: $\Pr(g_i|x_i)$.

This model is observationally equivalent. To see this, note that equation (5) implied that the optimal level of effort is common for students with the same x_i. The definition of y_i implies a closed-form solution: $y^*(x_i) = \frac{\Pr(g_i|x_i)}{1-\Pr(g_i|x_i)}$. We can therefore also write the flow utilities in the pure discrete choice model as $u^*(x_i) = u(x_i, y^*(x_i))$. Identification of $y^*(x_i)$ and $u^*(x_i)$ follows from the pure discrete case in Magnac and Thesmar (2002). This requires fixing β, normalizing the utility of one option and specifying the taste shock distribution.

Fixing $y^*(x_i)$ and $u^*(x_i)$ in counterfactuals implies no changes in effort. If identified, we could fix $c(x_i)$ and $C^0(x_i)$ instead. Appendix A.6 further discusses the differences in assumptions and implications of both models and section 6 quantifies the differences for the current application.

2.3.2 Simplified effort model

With the identified objects from the pure discrete model ($\Psi_1(x_i)$, $u^*(x_i)$, and $y^*(x_i)$), we can proceed to the identification of fixed and marginal costs of effort. The FOC identifies marginal costs from the marginal benefits at the optimal level of effort. Rearrange (4) and evaluate at $y_i = y^*(x_i)$ such that marginal costs can be written as a function of the identified objects:

$$c(x_i) = \beta \left(\frac{\ln\left(1 + \exp \Psi_1(x_i)\right)}{(1 + y^*(x_i))^2} \right). \tag{6}$$

To identify fixed costs, substitute $y_i^* = y^*(x_i)$ into the utility function of high school (1):

$$C^0(x_i) = -u^*(x_i) - c(x_i)y^*(x_i). \tag{7}$$

Intuitively, we are exploiting data on performance outcomes in a more structural sense than in a pure discrete choice model. If two students have the same future value $\Psi_1(x_i)$ but different state transitions (i.e. different $y^*(x_i)$), it has to be rationalized by differences in marginal costs. Students who make different choices in period 1 and have the same future values and state transitions provide the variation to identify the fixed costs.

The crucial assumption is that utility is linear in y_i (see (1)). This assures that fixed and marginal costs are policy-invariant functions of the state variables. Note that this holds for the specific definition imposed on y_i, i.e. about how the performance outcome depends on a choice variable that linearly enters the utility function.[3] The linear structure defines the transformations of utility that lead to policy-invariant functions. In this context, y_i is chosen to be the odds of obtaining a degree. This means that the linearity assumption should be interpreted as an assumption that the marginal cost of improving the odds for a given realization of the state variables does not change in counterfactuals.

2.3.3 General effort model

Since the degree of under-identification is the same as in a pure dynamic discrete choice model, it is straightforward to generalize the model to multiple periods t and alternatives j, as well as unobserved types ν_i. A pure discrete choice model would then identify flow utilities $u^*_j(x_{it}, \nu_i)$ and a transformation of performance data $y^*_j(x_{it}, \nu_i)$ (Hu and Shum, 2012; Kasahara and Shimotsu, 2009; Magnac and Thesmar, 2002). As

[3] This is different from a pure discrete (or a discrete/continuous) model that would include an observable measure of effort e_i in the model as e_i is data and cannot be specified by the researcher. Note also that by having the researcher specify y_i, x_i does not enter the probability to obtain a high school degree as we can always define a new $y_i^{'}$ such that $\phi(y_i') \equiv \phi(x_i, y_i)$.

I show in Appendix A, I can again exploit Euler equations to identify marginal costs $c_j(x_{it}, \nu_i)$ and use the remaining variation in program choices to identify fixed costs $C_j^0(x_{it}, \nu_i)$. I also provide extensions beyond the application of this paper by allowing y_i to be a vector, or to have direct effects beyond the next period state variables.

We can account for both unobserved ability and effort choice because we do not attempt to separately identify ability from a measure of effort we could observe in some datasets (such as hours of study). The choice variable here is *effective* effort. By construction, this merges the impact that ability, hours of study, or any other variable known to the agent would have on performance. Performance outcomes in the data can still deviate from a prediction based on effort, but only through an unexpected shock. This is why we could derive effort from the observed data by integrating over the shocks when unobserved ability did not enter the model (see (5)). To allow for unobserved ability, we first need to identify how performance depends on it, which is typically done in a pure discrete choice model too.

3 Institutional background and data

This section describes the institutional context in Flanders (Belgium) and introduces the data. I make use of the LOSO dataset in which I follow a sample of $5,158$ students that started secondary education in 1990.[4] Students were followed during high school and therefore the data contains many individual characteristics, choices, performance outcomes, and test scores. Afterward, they responded to surveys that reveal information about their higher education career. Details about the data and the context are discussed in Appendix B.

[4]The LOSO data were collected by Jan Van Damme (KU Leuven) and financed by the Flemish Ministry of Education and Training, on the initiative of the Flemish Minister of Education. Note that throughout the paper I discuss the data for this sample of 5,158 students, which covers 80% of the original sample. In Appendix B I discuss why some observations were dropped.

3.1 Study programs and observed characteristics

After finishing six grades in elementary school, students enroll in the high school of their choice in grade 7, the calendar year most of them become 12 years old. After obtaining a high school degree, they can enroll in higher education.

In full-time education, they choose between different high school programs, grouped into tracks that differ in their academic level. The curriculum of the academic track provides general education and prepares for higher education. The middle track prepares students for different outcomes. I follow Cockx et al. (2019) and distinguish between a track preparing mainly for higher education (middle-theoretical), and a track that prepares for the labor market (middle-practical). Students can also choose the vocational track, preparing them for occupations that do not require a college degree. Within tracks, I aggregate the variety of study programs in line with the most important differences in enrollment and success in higher education (Declercq and Verboven, 2015). The academic track includes classical languages, intensive math, intensive math + classical languages, and other. The middle-theoretical track has intensive math and other.

A student graduates from high school after a successful year in grade 12, except for the vocational track where completing a 13th grade is required.[5] Compulsory education laws require students to pursue education until June 30th of the year they reach the age of 18. From the age of 15, they can also decide to leave full-time education and start a part-time program in which work and schooling can be combined.

Table 1 provides an overview of the different programs students graduate from and how it differs by observable characteristics. The academic track is the most popular (38%), followed by the vocational track (19%), middle-theoretical track (16%) and the middle-practical track (12%). The remaining 15% dropped out. These groups are very different in terms of initial characteristics. I include cognitive ability (language

[5]Leaving in the vocational track after grade 12 is not considered dropout as students still obtain a certificate that is valued by employers.

and math), gender, and socioeconomic status (SES). The latter is defined as a dummy equal to one if at least one of the parents has completed higher education. Both SES and ability follow the track hierarchy. Female students are more likely to attend the academic track and less likely to drop out. There is also substantial variation in travel times to different programs and over time (Appendix Table J1), and students are more likely to attend a high school nearby (Appendix Table J2) or a college nearby (Appendix Table J5).

Table 1: Student background by high school program

	Number of students		Average characteristics			
	Count	(%)	Male	Language ability	Math ability	High SES
All	5158	(100.0)	0.50	0.00	0.00	0.28
By final study program						
Academic track	1974	(38.3)	0.40	0.71	0.64	0.49
clas+math	261	(5.1)	0.46	1.15	1.05	0.63
clas	315	(6.1)	0.37	0.94	0.68	0.58
math	683	(13.2)	0.49	0.74	0.75	0.51
other	715	(13.9)	0.32	0.41	0.36	0.38
Middle-Theoretical track	818	(15.9)	0.53	0.11	0.19	0.22
math	125	(2.4)	0.70	0.32	0.47	0.30
other	693	(13.4)	0.50	0.07	0.14	0.21
Middle-Practical track	611	(11.8)	0.51	-0.06	-0.02	0.22
Vocational track	1002	(19.4)	0.51	-0.76	-0.75	0.10
Dropout	753	(14.6)	0.67	-0.92	-0.86	0.07

NOTE.– Ability measured using IRT score on tests at the 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. 39.2% of students in the vocational track completed a thirteenth grade, 57.2% of students that dropped out chose part-time drop out first.

Similar to Declercq and Verboven (2015, 2018), I categorize higher education in Flanders into three levels: professional college, academic college, and university.

Additionally, I differentiate between the five university campuses and aggregate majors, distinguishing between STEM and non-STEM. Importantly, tracks do not restrict higher education options because colleges accept all students with a high school diploma with almost no restrictions on their major (Declercq and Verboven, 2018). Nevertheless, tracks of higher academic level strongly predict enrollment and graduation from them (Appendix Table J3 and J4).

3.2 Performance and the tracking policy

Students frequently finish in a different program than the one they entered. Initially, all programs are available, but moving up to a more academically oriented track is uncommon (see section 4.1 for details about the choice set). Consequently, many students begin with a more demanding program. While 63% of students start in an academic track, only 38% graduate from it (Appendix Figure J1).

Moves between programs are not always voluntary. Teachers uphold quality standards for each program by issuing a certificate based on the student's performance. An A-certificate is the best outcome, allowing students to advance to the next grade. However, if a student fails important courses, teachers may award a B- or C-certificate instead. A C-certificate indicates a failure in too many critical courses, requiring the student to repeat the grade. A B-certificate means that the student has failed some important courses and can only advance to the next grade in certain programs or repeat the grade. Typically, a B-certificate prevents a student from going to the next grade in their current track, although it may also exclude elective courses (see Appendix Table J7).[6]

While the yearly occurrence of B- and C-certificates is low (7.1% and 6.6% respec-

[6]Track switching and grade repetition are often handled differently in early tracking systems, yet similar trade-offs arise. The French-speaking Community of Belgium uses a similar system of certificates. In the Netherlands, schools have their own policies regarding track transitions and grade repetition, often in the form of grade requirements, combined with the judgment by teachers. In Germany and Austria, track revisions are less gradual, with mainly switches between middle and high school, and also more upward mobility (Dustmann et al., 2017; Schneeweis and Zweimüller, 2014).

tively), many students receive at least one of them during high school: 35% receive a B-certificate and 30% a C-certificate. If students don't leave full-time education, a C-certificate necessitates grade retention, and one in four students with a B-certificate repeats a grade, resulting in substantial grade retention. 32% of students leave high school with at least one year of study delay. These students are 22 % points less likely to enroll and 24 % points less likely to graduate from higher education compared to students who were not retained. This disparity is partly due to higher dropout rates in high school.[7]

4 Application of the model

This section extends the model of section 2 to the policy context discussed in section 3. i still refers to a student, t is a school year, $j = 1, ..., J$ are mutually exclusive study programs and $j = 0$ is an outside option: not attending school. Students make two yearly decisions: their study program (d_{it}) and their effort (y_{it}).

4.1 Choice set

Students enter high school the calendar year in which they become of age age_i^0 and choose a study program $d_{i1} = j$. Each study program belongs to one of four tracks: academic (*acad*), middle-theoretical (*midt*), middle-practical (*midp*), and vocational (*voc*). Within the academic track, they can choose to have intensive math courses (*math*), and/or classical languages (*clas*). A *math* option is also available in *midt*. Tracks are available throughout secondary education, i.e. grades 7 to 12 (13 in *voc*), denoted $grade_{it}$. The *clas* option is available from the start and the *math* options start in grade 9. A part-time vocational program (*part*) is available from the age of 15 and does not have a grade structure.

[7] Appendix Table J6 provides on overview of certificates by track and grade. Appendix Table J8 summarizes the number of students that obtain a B- or C-certificate or accumulate study delay. It then compares their educational outcomes with that of the average student.

The choice set of students (Φ_{it}) is restricted. First, students can never upgrade tracks according to the following hierarchy: $acad > midt > midp > voc > part$, except for the first two grades in which mobility between *acad*, *midt* and *midp* is allowed. Second, *math* and *clas* need to be chosen from the first year they are available. Finally, after grade 11, students must stay in the same program. Students progress by obtaining a certificate at the end of the year. The flexibility of a B-certificate affects the choice set differently. Consequently, I create new variables to capture which programs students are allowed to enter. After each year t, students obtain a vector of performance outcomes $g_{it+1} \in G$, with $g_{it+1} = (g^{track}_{it+1}, g^{math}_{it+1}, g^{clas}_{it+1})$. The main performance outcome is $g^{track}_{it+1} \in \{0, 1, 2, 3, 4\}$. The lowest value (0) does not allow any track in the next grade. Each increase corresponds to a track of higher academic level being available. $g^{math}_{it+1} \in \{0, 1, 2\}$ indicates if a student can go to *math*, with $g^{math}_{it+1} = 1$ when *math* is allowed in track *midt* ($g^{math}_{it+1} = 1$), $g^{math}_{it+1} = 2$ if it is allowed in both *midt* and *acad*. $g^{clas}_{it+1} \in \{0, 1\}$ indicates if a student can go to *clas*. In the last grade, the highest performance outcome of the track denotes graduation. A program-specific degree is obtained, denoted by a dummy $degree^{j}_{it}$, summarized in a vector $degree_{it} = (degree^{1}_{it}, ..., degree^{J}_{it})$. Students graduate after grade 13 in *voc* and grade 12 in other tracks. As finishing the 12th grade in *voc* is valued by employers, I also include a dummy: $degree^{voc_{12}}_{it}$. I assume students can no longer continue high school in $t = 10$, thereby allowing for up to three years of study delay.[8]

High school graduates can decide to enroll in different higher education options or leave school. I classify higher education into three levels (professional college, academic college, and university), and differentiate between STEM and non-STEM

[8]There are a few exceptions to these rules. First, it is allowed to switch from *acad* without extra math to *midt* with extra math in a later grade. Second, students can enroll in grade 8 of the vocational track without having succeeded grade 7. Therefore, the lowest performance outcome is a B-certificate and the effort costs of students in grade 7 of the vocational track is captured by a fixed component only. Note that the rules described in this section are not always legally binding, yet school often advertise them as binding. Cockx et al. (2019) apply a similar set of rules. In Appendix B I discuss more details and show that only a small number of observations have to be dropped because they are inconsistent with this description.

majors, as well as location. Once students turn 18, they can choose to leave school without a degree, and I assume this decision is permanent. As in section 2, I close the model at the college entrance, but I also predict college graduation as a function of individual characteristics and (endogenous) high school outcomes to evaluate the impact of counterfactuals in high school.[9]

4.2 State space

Each year t, students use three types of information to make their decisions. First, there is the information the student and econometrician share:

$$x_{it} = (d_{it-1}, g_{it}, age_i^0, grade_{it-1}, t, delay_{it}, degree_{it}, S_i).$$

To construct the choice set, we need information on $d_{it-1}, g_{it}, age_i^0, grade_{it-1}$ and t. Additionally, d_{it-1} allows us to take into account switching costs and g_{it} reveals if grade repetition is required for each program in the choice set. They also take into account past grade retention through study delay: $delay_{it} = \sum_{\tau<t} I(grade_{i\tau} = grade_{i\tau-1})$, with $I()$ the indicator function. $delay_{it}$, as well as the type of high school degree ($degree_{it}$) will be used to link high school outcomes with college enrollment and graduation. Finally, S_i is the vector of observed student characteristics.

Second, students have information about their type, unobserved to the econometrician: ν_i, a vector of dummy variables for each type.

Finally, there are idiosyncratic taste shocks for different programs: $\varepsilon_{it} = (\varepsilon_{i1t}, ..., \varepsilon_{iJt})$. This captures the unobservables students learn about at the start of period t.

4.3 End-of-year performance

Track restrictions

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

The performance measure g_{it+1}^{track} is the result of effort y_{it} and a logistically distributed shock η_{it+1}^{track}:

$$g_{it+1}^{track} = \tilde{g} \text{ if } \bar{\eta}_{jr}^{\tilde{g}} < \ln y_{it} + \eta_{it+1}^{track} \leq \bar{\eta}_{jr}^{\tilde{g}+1} \tag{8}$$

where $\bar{\eta}_{it}^{\tilde{g}}$ denotes the threshold to obtain at least $\tilde{g}$. It is allowed to differ through the program and grade a student is attending at time t. At time t, students know y_{it}, but not g_{it+1}^{track} because of the shock η_{it+1}^{track}. They know the probabilities in a given program ($d_{it} = j$) and grade ($grade_{it} = r$):

$$\Pr(g_{it+1}^{track} = \tilde{g}|y_{it}, d_{it} = j, grade_{it} = r) = F(\ln y_{it} - \bar{\eta}_{jr}^{\tilde{g}}) - F(\ln y_{it} - \bar{\eta}_{jr}^{\tilde{g}+1}) \tag{9}$$

with $F(a) = \frac{\exp(a)}{1+\exp(a)}$ the cumulative distribution function of the performance outcome. I set $\bar{\eta}_{jr}^{0} = -\infty$ and $\bar{\eta}_{jr}^{5} = +\infty$ such that the probabilities add up to 1, and I normalize $\bar{\eta}_{jr}^{1} = 0$.

Effort can be interpreted here as the odds of avoiding the lowest outcome (= no track allowed in the next grade):

$$y_{it} = \frac{1 - \Pr(g_{it+1}^{track} = 0|y_{it}, d_{it}, grade_{it})}{\Pr(g_{it+1}^{track} = 0|y_{it}, d_{it}, grade_{it})}. \tag{10}$$

This is a natural extension of the binary case we discussed in section 2. By choosing these odds, students can change the probability of each realization. If y_{it} is close to zero, they are likely to obtain the worse outcome, while high values yield the best outcome. Several thresholds are not estimated but determined by the restrictions discussed in subsection 4.1.

Course restrictions

Course restrictions are also modeled as an ordered logit (conditional on g_{it+1}^{track}) to predict g_{it+1}^{math} and g_{it+1}^{clas}. I specify course-specific indexes:

$$\begin{gathered}\alpha_y^{math} \ln y_{it} + S_i' \alpha_S^{math} + \nu_i' \alpha_\nu^{math} + \eta_{it+1}^{math}, \\ \alpha_y^{clas} \ln y_{it} + S_i' \alpha_S^{clas} + \nu_i' \alpha_\nu^{clas} + \eta_{it+1}^{clas}.\end{gathered} \tag{11}$$

with η_{it+1}^{math} and η_{it+1}^{clas} logistically distributed shocks and α_y^{math} and α_y^{clas} measuring how much of the effort matters for each elective course. The logit shocks η_{it+1}^{math} and η_{it+1}^{clas} are assumed to be independent of η_{it+1}^{track} but dependence between outcomes is captured by taking into account the outcome g_{it+1}^{track} to influence individual-specific threshold levels.[10] I also allow for comparative advantages by estimating the effect of student characteristics that can be observed (S_i) or unobserved (ν_i). In counterfactuals, I will treat the parameters of this equation as policy-invariant.[11]

I define $\phi_{ijt}^{\bar{g}}(y_{it})$ as the joint probability of $\bar{g} = \{\bar{g}^{track}, \bar{g}^{clas}, \bar{g}^{math}\}$ which is the product of the three ordered logit probabilities, with dependence on i and t going entirely through (x_{it}, ν_i).

4.4 Study program

Students choose the program j and effort y_{it} that gives the highest expected lifetime utility. For a given program, it can be written as:

$$\begin{aligned} & v_j(x_{it}, \nu_i, y_{it}) + \varepsilon_{ijt} \\ & = u_j(x_{it}, \nu_i, y_{it}) + \beta \sum_{\bar{g} \in G} \phi_{ijt}^{\bar{g}}(y_{it}) \bar{V}(x_{it+1}(\bar{g}), \nu_i) + \varepsilon_{ijt} \text{ for } j \in se \end{aligned} \tag{12}$$

with $v_j(x_{it}, \nu_i, y_{it})$ the conditional value function of choosing program j and effort y_{it} at time t.[12] ε_{ijt} is an extreme value type 1 taste shock.

[10] If $g_{it+1}^{track} < 3$, then $g_{it+1}^{math} = g_{it+1}^{clas} = 0$. If $g_{it+1} = 3$, then $g_{it+1}^{math} \in \{0, 1\}$ and $g_{it+1}^{clas} = 0$.

[11] As shown in Appendix A.3, y_{it} could instead be treated as a vector at the cost of increasing the computational burden in counterfactuals.

[12] I do not let the state vector of other students affect the value functions of student i. This implies that program-specific costs and benefits do not change in counterfactuals. The model then generates

The first term is the flow utility of schooling, $u_j(x_{it}, \nu_i, y_{it})$. The second term is the expected value of the future, discounted by $\beta \in (0, 1)$. This depends on the ex-ante value functions $\bar{V}(x_{it+1}, \nu_i)$, i.e. the value functions integrated over the future iid shocks. As in Rust (1987), this implies that students do not know future realizations of taste shocks, but they know the distribution. The performance vector g is the only stochastic element in x. Integrating over future states is therefore equivalent to writing a weighted sum over potential outcomes in the set G, with the joint probability of the performance outcome as a weight.

As explained in section 2, I do not use estimates of $u_j(x_{it}, \nu_i, y_{it})$ as a function of (x_{it}, ν_i) in counterfactual simulations as it would ignore potential changes in effort. Instead, I estimate effort costs: fixed cost $C_j^0(x_{it}, \nu_i)$, and marginal costs $c_j(x_{it}, \nu_i)$ with

$$u_j(x_{it}, \nu_i, y_{it}) = -C_j^0(x_{it}, \nu_i) - c_j(x_{it}, \nu_i) y_{it}. \tag{13}$$

The linearity assumption implies that the marginal costs of increasing the odds to avoid the lowest performance outcome do not change in counterfactuals. Note that effort (y_{it}) only affects the future indirectly through its impact on performance. In Appendix F, I show that this is of little concern in the current context by analyzing test score data during high school.

4.5 Closing and solving the model

I assume that leaving secondary education is a terminal action. They either leave the education system or (if they obtained a high school degree) enter higher education. I close the model at the enrollment stage of higher education to avoid making assumptions about how students expect wages and college performance to evolve. Let the

the partial equilibrium effects of a policy, which are appropriate for simulations that mainly affect how students are tracked in the current system, without large changes in the average composition of each program.

conditional value functions for options after high school take the following form:

$$v_j(x_{it}, \nu_i, y_{it}) = degree'_{it}\mu^{\text{degree}} + \Psi_j^{HEE}(x_{it}, \nu_i) \text{ if } t = T_i^{SE} + 1 \tag{14}$$

with T_i^{SE} the last period student i spends in high school, μ^{degree} a vector of parameters and $\Psi_j^{HEE}(.)$ a function of the state variables. With a high school degree, j can be a higher education option (see subsection 4.1). All students have access to an outside option ($j = 0$) with $\Psi_0^{HEE} = 0$. This normalization implies that costs in high school should be interpreted as the one-period difference from the expected lifetime value of leaving high school without a degree. As this includes their wages, effort costs in the model also include these opportunity costs.[13]

By modeling higher education choices, we can account for the varying returns of different high school programs. For instance, a math-intensive program may prove beneficial in higher education, particularly for STEM majors, which will be reflected in a different value of $\Psi_j^{HEE}(x_{it}, \nu_i)$. Moreover, I incorporated heterogeneity in the costs of high school programs ($C_j^0(x_{it}, \nu_i)$ and $c_j(x_{it}, \nu_i)$). Allowing for such rich observed and unobserved heterogeneity is important to evaluate policy changes because the returns of affected students often differ from those of the average student (Carneiro et al., 2011).

Unlike a static model, normalizing the entire utility of the outside option in every state is not innocuous (Kalouptsidi et al., 2021). In this case, it would assume that students exert effort in school solely to have the chance to pursue higher education and not for other benefits resulting from a high school degree. Therefore, I also estimate that value such that $v_{i0t} = degree'_{it}\mu^{\text{degree}}$. As in Eckstein and Wolpin (1999), this can be identified from schooling choices in secondary education as students closer to graduation (higher expected performance or attending a higher grade) are less

[13] The outside option may not be available due to compulsory schooling laws, but this is not a problem for identification as we include a sufficient number of students in the sample that are old enough. It does imply that utility cannot vary completely by age, but we do consider variation by grade.

likely to drop out. Because such differences could also be explained by a flexible specification of fixed costs, I restrict the latter to only change linearly by grade and I exclude distance to college options.

As in section 2, the model can be solved backwards. Appendix C describes the full solution.

4.6 Graduation in higher education

I simultaneously estimate the parameters of a conditional logit model with $\Psi_j^{HED}(x_{it}, \nu_i)$ the estimated index that predicts graduation in each campus-level-major combination, conditional on student characteristics, high school program, study delay, and the higher education enrollment decision. This reduced-form approach is sufficient because the counterfactual simulations will only modify the high school system, not the higher education system. As a result, it will affect elements of x_{it}: $degree_{it}$ and $delay_{it}$, but not the mapping between high school and higher education outcomes: $\Psi_j^{HEE}(.)$ and $\Psi_j^{HED}(.)$. This approach is similar to that used in dynamic treatment effect models (Heckman et al., 2016).

4.7 Identification, ability bias and unobserved types

A first requirement to identify the model is to recover CCPs and state transitions as functions of the observed state variable x_{it} and the unobserved type ν_i. If ν_i would be observed, we could use the observed choices and outcomes for each realization of (x_{it}, ν_i). Magnac and Thesmar (2002) then show that we need to normalize the utility of a reference alternative, specify the discount factor β and the distribution of ε_{ijt} to identify the flow utility in the current policy context ($u_j^*(x_{it}, \nu_i) \equiv u_j(x_{it}, \nu_i, y_j^*(x_{it}, \nu_i))$). The identification of the indexes that predict higher education enrollment and graduation is simpler because we do not need to separately identify flow utility from the entire impact of (x_{it}, ν_i) (French and Taber, 2011). In section 2 and Appendix A I show that once we recover flow utility and state transitions, we can identify fixed costs $C_j^0(x_{it}, \nu_i)$ and marginal costs $c_j(x_{it}, \nu_i)$ by imposing a FOC. I set $\beta = 0.9$ and show robustness in Appendix F.

Since unobserved heterogeneity affects choices and outcomes in high school and afterwards, it creates a classic ability bias problem to assess the impact of high school outcomes (study delay and study program) in higher education. Exclusion restrictions identify this without needing the full structure of the model (Heckman and Navarro, 2007). In particular, I assume that travel time to high school (varying by program and grade) influences decisions and outcomes during high school but has no direct effect afterwards. Note that this variation also helps to identify the effects of study delay because students who enter higher education might have obtained a B-certificate in the past, giving them the choice to accumulate study delay.[14]

In the model, I capture unobserved heterogeneity by allowing for two types (as in Arcidiacono (2005); Declercq and Verboven (2018)). Any (causal) claim we make depends on our ability to approximate the heterogeneity in the population by the observable characteristics, the two unobserved types and the functional form assumptions we make (see next subsection), but this choice has several benefits. It allows for the use of the CCP estimator, which yields large computational advantages (Arcidiacono and Miller, 2011). Moreover, it allows for a flexible correlation structure of unobserved heterogeneity in utilities and outcomes. Hu and Shum (2012) prove the identification of a non-stationary first-order Markovian model for CCPs and state transitions at time t using data from $t+1$, t, $t-1$, $t-2$, and $t-3$. They allow for a single unobserved trait (potentially transitioning over time). Because high school takes six years to complete and we add two stages after high school, this shows that no

[14] As we argue in De Groote and Declercq (2021), this context lends itself to the use of this instrument as students have many school options available to them and parents are therefore not expected to take this into account in their location decisions. Importantly, free school choice is protected by the Belgian constitution and prevents schools from cream-skimming or prioritizing students of the same neighborhood. Note that Heckman and Navarro (2007) do not require an exclusion restriction. One can for example also use an identification at infinity strategy (Abbring, 2010; Heckman et al., 2016). In Appendix F I show that the main results are robust to adding measures of travel time to high school to the equations that predict higher education outcomes.

further structure is needed to identify CCPs and state transitions at the end of high school and the enrollment stage of higher education. Furthermore, the dependence on unobserved heterogeneity of CCPs and state transitions only goes through a single unobserved factor: their type. This restriction allows for the identification of a broad set of distributional treatment effects (Carneiro et al., 2003; Heckman et al., 2016), which is important for the channels that drive the counterfactuals: the causal impact of tracks and grade retention. Adding noisy measures of unobservables outside the model is not necessary (Freyberger, 2018), but can help identification. I do this in Appendix F to show the robustness of results.

To investigate the impact of observing measures of ability and adding unobserved types, I provide a sensitivity analysis in Appendix F. This yields three main conclusions. First, including the two unobserved types or the rich measures of ability decreases the negative impact on college graduation from the downgrade policy and increases the negative impact of the repeat policy. Second, the total impact is mainly driven by observed ability for the repeat policy, but by unobserved types for the downgrade policy. Finally, adding a third unobserved type changes little to the main results.

4.8 Estimation

I summarize the estimation algorithm and parametric assumptions in the main text and discuss it in detail in Appendix D. It is an application of the two-stage CCP estimator of Arcidiacono and Miller (2011):

Stage 1: estimate type distribution and reduced forms

Step 1: initial types

Assume there are two unobserved types, assign each student a random probability and use it as weights in what follows.

Step 2: higher education estimates

Impose functional forms for $\Psi_j^{HEE}(.)$ and $\Psi_j^{HED}(.)$ and estimate them as parameters of a conditional logit, using maximum likelihood. Importantly, I include characteristics of students (observed S_i and unobserved ν_i) as well as the high school track at graduation ($degree_{it}$) and the years of study delay ($delay_{it}$).

Step 3: reduced forms of high school data

Recover the optimal levels of effort and the performance thresholds by estimating an ordered logit model with index $\ln y_j^*(x_{it}, \nu_i)$. The index is specified as a flexible function of (x_{it}, ν_i). I allow for effects of student characteristics S_i and ν_i, that differ by track and elective course. The track effect can change linearly with the grade. I estimate the impact of travel time and the past track and elective course through d_{it-1}. Grade retention is captured by a stock variable ($delay_{it}$) and a flow variable (a dummy for repeating a grade) and I include the distances to different higher education options.

I also estimate an ordered logit to recover the parameters of the performance outcomes on elective courses in equation (11) and I follow Arcidiacono et al. (2023) in obtaining predicted values of CCPs by estimating a flexible conditional logit using a similar index as I used to model performance.

Step 4: update types

Use the likelihood contributions of CCPs, performance and higher education outcomes by type to update the individual-level type probabilities using Bayes rule.

Repeat this until convergence of the joint likelihood.

Stage 2: estimate cost parameters

Use the logit probabilities with the CCP representation of the conditional value functions to estimate the value of a degree μ^{degree} and fixed costs $C_j^0(.)$ using maximum likelihood, with type probabilities as weights. I assume fixed costs include a program-specific constant, travel time, and switching costs between tracks and specializations through elective courses. They also differ because of individual characteristics S_i and ν_i through an effect that is allowed to change linearly in the level of the track and by elective course. Finally, grade retention enters through accumulated study delay and a dummy for repeating the grade. These effects are also allowed to change linearly with the academic level.

Finally, marginal costs $c_j(.)$ can be recovered from the FOC at the optimal levels of effort $(y_j^*(x_{it}, \nu_i))$, without imposing additional structure:

$$c_j(x_{it}, \nu_i) = \beta \sum_{\bar{g}} \frac{\partial \phi_{ijt}^{\bar{g}}(y_{it})}{\partial y_{it}} \bar{V}(x_{it+1}(\bar{g}), \nu_i) \text{ if } y_{it} = y_j^*(x_{it}, \nu_i). \tag{15}$$

Note that $y_j^*(x_{it}, \nu_i)$ was identified in stage 1, step 3. $\frac{\partial \phi_{ijt}^{\bar{g}}(y_{it})}{\partial y_{it}}$ can be calculated using the ordered logit probabilities. The CCP representation is used to write $\bar{V}(x_{it+1}(\bar{g}), \nu_i)$.

Standard errors are obtained using a bootstrap procedure.[15]

5 Estimation results

This section discusses the effort cost estimates and the estimates of higher education outcomes. The model fit and details about the simulations are explained in Appendix E.

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

5.1 Effort costs

Table 2 shows the impact of student characteristics on effort costs, while Table 3 displays the impact of study delay and program switches. Appendix Table J11 presents the estimates of travel time, grade and track intercepts. Table J12 the interactions of student characteristics with elective courses and Table J16 the intrinsic value of a high school degree.

Table 2: Costs of schooling: student characteristics and academic level

	Fixed costs				Log of marginal costs			
	Baseline effect		× academic level		Baseline effect		× academic level	
Language ability	7.5	(5.5)	-36.3	(5.1)	-0.64	(0.08)	-0.10	(0.09)
Math ability	1.1	(5.1)	-23.2	(4.5)	-0.22	(0.06)	-0.37	(0.05)
High SES	-19.2	(15.6)	-18.8	(5.8)	-0.68	(0.22)	0.02	(0.10)
Male	-19.2	(9.7)	17.7	(4.4)	0.74	(0.11)	0.10	(0.06)
Type 2	-41.8	(16.4)	85.6	(11.6)	3.27	(0.37)	-0.37	(0.16)

NOTE. – Estimates of a sample of 5,158 students or 33,239 student-year observations. Scale = minutes of daily travel time. The marginal costs in the model are a flexible function of state variables, this table summarizes them by regressing their logarithmic transformation on the same variables that enter the fixed costs. Ability measured in standard deviations. High SES= at least one parent has higher education degree. Type 2 = dummy equal to one if the student belongs to unobserved type 2 instead of 1. Academic level = academic level of high school track (0: vocational, 1: middle-practical, 2: middle-theoretical, 3: academic). Bootstrap standard errors in parentheses.

Regarding fixed costs, the functional form assumptions are consistent with the tables provided and the parameters are scaled in daily minutes of travel time.[16] The marginal costs are a function of probabilities in the data and other parameters of the model (see Appendix D). For interpretational purposes, I perform an OLS regression on the logarithmic transformation of the estimated marginal costs, using the same

[16] I do this by dividing by the travel time parameter in utility, which is precisely estimated: -0.00597, with standard error 0.00054. This can be used to obtain the static elasticity (i.e. keeping future utility fixed) of choosing an alternative with respect to its travel time (Train, 2009, pp59): $-0.00597 \times time_{ijt} \times (1 - \tilde{p_{ijt}})$, with $\tilde{p_{ijt}}$ the predicted choice probabilities. In period 1 (when everyone has the same choice set), this is on average -0.17.

structure as for fixed costs.[17]

Cognitive ability has large effects. The fixed costs of the benchmark vocational program is not affected in a statistically significant way, but there are large differences between tracks. Specifically, a one standard deviation increase in language ability leads to a decrease in fixed costs equivalent to an additional 36 minutes of travel time to school for each unit increase in the academic level. Such an increase in math ability decreases this cost by 23 minutes. Marginal costs are affected too. An increase of 1% of a standard deviation in language ability leads to a decrease of 0.6% in marginal costs in the vocational track (level=0), and 0.9% in the academic track (level=3). For math ability we find a 0.2% decrease in marginal costs in the vocational track, but a much larger 1.3% decrease in the academic track.[18]

Conditional on cognitive ability, high SES students are still more favorable towards programs of higher academic level, decreasing fixed costs by 19 minutes for each step increase in the level of the track. This could reflect intrinsic preferences for being in a higher track or parents encouraging their children to choose tracks that are more in line with their own education. This group also has substantially lower marginal costs. In all tracks, a high SES student pays only half ($\exp(-0.68)$) the cost of a low SES student for the same increase in effort. This reflects a different study environment or differences in the intrinsic motivation to study. Male students have a 19-minute lower fixed cost to attend the vocational program, but the gender effect reverses for the most academic programs. The marginal cost estimates suggest that they face greater difficulty achieving good performance outcomes.

Observable characteristics do not fully explain the persistent heterogeneity in the data. Appendix Table J13 reveals that 30% of students belong to type 1 and 70% to

[17] The main policy simulations in this paper ("downgrade" and "repeat"), as well as a simulation where I adjust the age at which education is no longer compulsory, do not make use of this approximation. In other simulations I do use this approximation because it is more convenient and transparent to adjust them for the counterfactual.

[18] I also estimated a model where academic level is proxied by the hours of academic courses (which varies over both tracks and grades) and obtain similar results.

Table 3: Costs of schooling: repeating and switching

	Fixed costs				Log of marginal costs			
	Baseline effect		× academic level		Baseline effect		× academic level	
Repeat	274.7	(33.2)	79.9	(13.1)	0.24	(0.19)	-0.46	(0.12)
Study delay	-9.9	(6.9)	10.0	(4.9)	0.64	(0.10)	0.11	(0.06)
Downgrade	169.1	(21.9)			0.17	(0.09)		
Upgrade	325.7	(41.6)			0.26	(0.18)		
Stay in clas	-16.1	(14.7)			0.16	(0.43)		
Stay in math	-203.5	(26.1)			0.65	(0.28)		

NOTE. – Estimates of a sample of 5,158 students or 33,239 student-year observations. Scale = minutes of daily travel time. The marginal costs in the model are a flexible function of state variables, this table summarizes them by regressing their logarithmic transformation on the same variables that enter the fixed costs. Academic level = academic level of high school track (0: vocational, 1: middle-practical, 2: middle-theoretical, 3: academic). Downgrade: switch to a lower academic level. Upgrade: switch to a higher academic level. Clas= classical languages included. Math= intensive math. Bootstrap standard errors in parentheses.

type 2. Type 1 students excel in high academic tracks and face much lower fixed costs, equivalent to 86 minutes of daily travel time for each step increase. They also have much lower marginal costs, paying only 4% to 12% of the costs of type 2 students.[19]

Table 3 shows how track choices and grade retention impact costs during high school.

Study delay, i.e. past grade retention, increases marginal costs, potentially due to demotivation. Repeating a grade in programs of high academic level decreases marginal costs, possibly due to familiarity with course material. However, the fixed cost estimates suggest that students strongly dislike repeating a grade.

Finally, students do not like to switch programs. Both down- and upgrading is associated with much higher fixed costs. Note that upward mobility is only allowed in the first two high school grades, but also here we see larger costs than for downward switching. This could reflect the difference in the way schools advertise this possibility

[19] In the benchmark (=vocational) track this is $\exp(-3.27) = 4\%$, in the academic track it is $\exp(-3.27 + 3 \times 0.37) = 12\%$.

as downward mobility is common, while upward mobility rarely happens.[20] Both types of mobility increase marginal costs, but the estimates are not precise. Note that this does not imply that a switch cannot reduce costs as it could lead to a better match between ability and track.

5.2 Higher education

Estimates for higher education can be found in Appendix Tables J17, J18, and J19. Distance and student characteristics, including the unobserved types, impact college enrollment and graduation. To facilitate the interpretation of the effect of high school study programs, I calculate the Average Treatment effects on the Treated (ATT) and compare this to the difference in means in the data in Table 4.

This number is calculated as a "ceteris paribus" effect, i.e. keeping all other variables that were realized before or at the time of leaving secondary education fixed (see Appendix D.6). Similarly, I calculate the effect of one year of study delay by comparing outcomes for retained students in the counterfactual scenario where they would not have accumulated study delay.

Most estimates point in the same direction as a comparison of means, but to a much smaller extent. Graduating from the academic track (without classical languages or intensive math) increases college graduation by 20 % points compared to the middle-practical track. This shrinks to 9.5% points when compared to the middle-theoretical track. Elective courses mainly matter for the type of higher education but we also see increased graduation rates for students who had classical languages or intensive math.

Study delay decreases higher education enrollment by 5 % points and graduation by 12 % points.

[20] In Appendix E I show predictions of a model that encourages upward mobility.

Table 4: Higher education and high school outcomes: difference in means and ATTs

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

NOTE. – Effects on enrollment and degree completion after graduating from different high school programs, compared to graduating from the academic track without clas or math option, and the effects of one year of study delay, compared to 0. Average treatment effects on the treated (ATT) make use of the estimates of enrollment and graduation equations. ATTs are calculated using indexes, specified in Appendix D, for each individual at the realization of other variables. Effects on obtaining higher education degree are total effects, i.e. they also take into account effects through enrollment. Clas= classical languages included. Math= intensive math. Bootstrap standard errors in parentheses.

6 Counterfactual tracking policies

As explained in section 3, teachers determine if a student has the skills to move to the next grade in their program. 35% of students receive a B-certificate at some point during their secondary education, meaning their performance is insufficient to transition to the next grade unless they decide to switch to a different program. The status quo scenario presents these students with two options: repeat or downgrade. I compare this to two counterfactual scenarios that force each option:

Counterfactual 1: Repeat

Students have to repeat a grade when they obtain a B-certificate. It makes the system less flexible and allows us to quantify the importance of the current flexibility.

Counterfactual 2: Downgrade

Students have to switch to a different program when they obtain a B-certificate, without repeating the grade. It resembles a new policy in Flanders to reduce grade retention and allows us to quantify the importance of offering students the possibility to ignore the advice of teachers.

I first discuss the predicted effect of each policy.[21] I then discuss the role of parental background and other sources of heterogeneity in the current and alternative tracking contexts and I conclude with a discussion on optimal policy. Details about the calculation of welfare effects, as well as alternative policy simulations, can be found in Appendix E.

6.1 Policy impact

I first discuss the results in the model of effort choice and then show the difference if we would have used a pure discrete choice model instead.

6.1.1 Model with effort choice

Table 5 compares the outcomes of the two counterfactuals to the status quo scenario.

The "repeat" policy leads to worse outcomes, as it fails to significantly increase graduation rates from the academic track and increases dropout rates by 4% points,

[21] The downgrade policy is part of a reform in secondary education and is applied on cohorts that entered high school from September 2019 on. In the implementation of the policy, it is still possible in some cases to repeat the grade but only if students get their teachers' explicit permission (source: answer by the Flemish minister in parliament at 4 October 2018 on question 2410 in period 2017-2018).

a 28% increase in the total number. Moreover, the share of students with grade retention increases substantially (9 % points), causing enrollment and graduation rates in higher education to drop by 2 % points. The SES gap in college remains unaffected. Assuming an opportunity cost of \$10/hour, I find that student welfare decreases on average by \$2,140, mainly driven by the increase in fixed costs resulting from repeating grades. The increase in dropout rates also reduces the expected payoff after leaving high school, as these students are unable to enroll in college. The increase in grade retention and the decrease in college graduates are also expected to have large negative externalities that are not considered in this exercise.

The "downgrade" policy results in a decrease in grade retention rates by 10 % points and dropout rates by 1.6 % points. This comes with a cost in the short run. Students switch to programs of lower academic level, which decreases enrollment rates in higher education by 1.4 % points. Graduation rates only decrease by an insignificant 0.3 % points (or less than 1% of the total number), which can be explained by the negative effect of study delay on graduation. Again, the SES gap in college remains unaffected. The policy restricts the choice set of students, leading to an average welfare loss of \$1,020, despite a reduction in the fixed costs of \$480 and an increase in their expected payoff after high school of \$320.[22] The loss is explained by an increase in effort (variable costs increase by \$210), but also a loss of \$1,610 in taste shocks. The latter captures unobserved, time-varying preferences such as preferences for certain teachers or classmates.

[22]Since we close the model after high school, the utility of enrolling in college is the students' expected lifetime utility at the time they leave high school. These expectations can be biased. By simultaneously predicting higher education graduation, we can see that the negative impact of study delay is much larger for graduation than for enrollment (see Table 4). Similarly, the counterfactual impacts the college enrollment rate but it does not have a significant impact on the number of college graduates. If students ultimately care about graduation, rather than enrollment, it suggests that they might underestimate the negative consequences of study delay in the long run. It is therefore possible that the increase in the expected payoff is substantially smaller than the increase in the actual payoff.

6.1.2 Difference with the pure discrete choice model

To demonstrate the importance of allowing effort to be a choice variable, Table 6 compares the counterfactual predictions with those from a pure discrete choice model. Note that stage 1 of the estimator identifies the type distribution, the reduced forms of high school and the higher education estimates. These are the same in the pure discrete choice model. The only difference is that in stage 2 "fixed" costs now capture the entire flow utility and are therefore also re-estimated (i.e. there is no term $c_j(x_{it}, \nu_i)y_{it}$ in the utility function (13) anymore).

A model without effort choice leads to worse results in both counterfactuals. For example, the increase in study delay in the "repeat" policy is 11.6 % points instead of 9.5. The decrease in the "downgrade" policy is 9.2 % points instead of 9.8. For higher education, we would conclude that there is an important negative impact on higher education graduation from the "downgrade" policy (-1.1 % points), while the model with effort choice only estimates an insignificant and small effect of -0.3.

The difference in results can be attributed to the impact of the policy on study effort. Both counterfactuals decrease the value of a B-certificate. In a dynamic model with program choices, students can avoid this by choosing programs with higher success rates. With effort in the model, they could instead improve their success rate. Particularly students who are likely to receive a B-certificate and who dislike the new policy are expected to pay this cost.

In Appendix G I show that the new policies increase effort at the beginning of secondary education. The impact is most clear for the predicted number of bad performance outcomes in Table 6, and especially B-certificates. In the "repeat" policy, the decrease in the number of B-certificates is only 62% of the decrease in a model where students can adjust their study effort. In the "downgrade" policy it is only 25%. This has important implications. First, there is a smaller increase in study delay in the "repeat" policy and a stronger decrease in the "downgrade" policy. Second, more students are staying in academic programs. This increase in study effort is what explains the impact on variable costs in models with effort choice, but the impact on total student welfare is the same. The more favorable higher education outcomes compared to a pure discrete choice model are a result of the decrease in dropout, the increase in students graduating from academic programs, and the decrease in study delay.

6.2 Parental background and other initial conditions

Early tracking of students can worsen the impact of initial conditions, especially parental background, on educational choices. To understand this impact in the status quo and the two counterfactual scenarios, I run regressions on the main outcomes of interest from all the simulated data. This summarizes the ceteris paribus impact of each initial condition in the status quo and counterfactual scenarios. I then run additional simulations to decompose the effects of parental background on the most important outcomes.

The regression results can be found in the Appendix Table J20. I find that the impact of the "repeat" or "downgrade" policy does not vary much with student characteristics. However, the direct impact of initial conditions on outcomes is large. A decrease in language ability by 10% of a standard deviation makes a student 0.7 % points less likely to obtain study delay, 0.7 % points less likely to drop out, 1.5 % points more likely to graduate from college and derive $\$1,627$ more from the high school system. The impact of math ability is similar but less important for study delay. The impact of high SES and being female is identical in sign and similar in magnitude to a standard deviation increase in ability, while the differences between the two unobserved types are about twice this size.

The results of giving low SES students some of the high SES parameters or initial conditions can be found in Appendix H. They can be summarized as follows: (1) the main reason behind the SES gap in college is coming from ability before high school entry, (2) encouraging low SES students to opt for more academic programs by providing them with the fixed costs of high SES students reduces the SES gap but increases study delay. The latter can be avoided by bringing their marginal costs also down to the level of high SES students (e.g. due to a potentially better study environment).

6.3 Conclusions for optimal policy

We can derive two main conclusions to inform optimal policy.

First, it's better to encourage underperforming students to switch to lower tracks instead of repeating a grade. The lack of a downgrade option in the "repeat" policy does not help them to succeed in a more academic track and comes at a high cost in terms of welfare. However, this does not mean that the status quo policy of allowing students to choose is optimal. The "downgrade" policy shows that reducing study delay and dropout is possible without negative consequences on higher education graduation by forcing them to downgrade. While such a reduction in the choice set inevitably leads to a decrease in estimated welfare, it can be argued that the "downgrade" policy is still beneficial for society.[23] First, OECD (2012; 2013) estimates show that the per capita loss of students ($\$1,020$) is close to the government saving by financing fewer years of schooling ($950). Moreover, gains from a year in taxes yield an additional $\$1,960$. The decrease in dropout and the improved efficiency of higher education (which is 90% government-funded) would generate even larger returns. There are also reasons to believe that students gain directly from the "downgrade" policy in a way we did not account for in the model. Assuming students can be represented by rational agents rules out ex-ante mistakes. The increasing evidence in the context of educational decisions shows that regulating choices could also improve student

[23] I also investigated alternative policies that do not restrict students' choice set: the impact of lowering the age of compulsory education, and allowing for upward mobility, the results can be found in Appendix E. Lowering the age of compulsory education reduces grade retention because it gives underperforming students an alternative way out. However, it also increases drop out substantially. Unrestricted upward mobility reduces graduation from the vocational track, but it also shifts many students into programs in which they underperform, leading to negative outcomes resulting from the increase in study delay.

welfare (Bhargava and Loewenstein, 2015; Koch et al., 2015; Lavecchia et al., 2016). In Appendix I, I discuss that counterfactual choice probabilities are likely robust to common sources of mistakes, but that welfare estimates can be biased, implying that students would be better off with the "downgrade" policy than predicted.

Second, the initial conditions of students when they enter high school have a big impact on outcomes in all tracking policies considered here. Parental background matters, mostly (but not only) because it is correlated with ability before entering high school.

In summary, high schools should focus on efficiently fostering skills that students have already acquired and avoid having them repeat grades in pursuit of an academic curriculum, which is both ineffective and costly. This is consistent with the research on the benefits of early childhood education and its dynamic complementarities (Cunha and Heckman, 2009; Cunha et al., 2010; Heckman and Mosso, 2014). The cost savings resulting from reducing study delay could be invested in improving initial conditions and thereby improving student outcomes. Appendix E.4 provides back-of-the-envelope calculations of the expected effects.

Table 5: Counterfactual tracking policies: implications B-certificate

	Status quo prediction	Changes after repeat policy		Changes after downgrade policy	
Educational outcomes (%):					
Final high school track					
Academic	40.02	0.17	(0.30)	-1.13	(0.32)
Middle-theoretical	16.10	-0.96	(0.29)	-1.52	(0.26)
Middle-practical	8.14	-1.13	(0.24)	-0.77	(0.26)
Vocational	21.57	-2.01	(0.32)	5.03	(0.29)
Dropout	14.17	3.94	(0.33)	-1.61	(0.25)
At least 1 B-certificate	37.53	-10.23	(0.71)	-3.49	(0.34)
At least 1 C-certificate	30.69	-0.23	(0.33)	-2.15	(0.24)
At least 1 year of study delay	33.22	9.48	(0.57)	-9.82	(0.55)
Higher education					
Enrollment	58.15	-1.76	(0.24)	-1.40	(0.21)
Graduation	44.25	-1.70	(0.22)	-0.30	(0.18)
SES gap at graduation	39.73	0.06	(0.33)	0.11	(0.25)
Student welfare ($1000):					
Fixed costs (-)		0.85	(0.12)	-0.48	(0.10)
Variable costs (-)		0.49	(0.08)	0.21	(0.03)
Expected payoff after high school (+)		-0.65	(0.10)	0.32	(0.07)
Taste shocks (+)		-0.15	(0.10)	-1.61	(0.19)
Total		-2.14	(0.26)	-1.02	(0.14)

NOTE. – Predictions from the dynamic effort model. C-certificate: repeat grade. B-certificate = students acquired skills to proceed to next grade but only if they downgrade, i.e. switch to track of lower academic level or drop an elective course. Status quo = students can choose to downgrade or repeat grade after obtaining a B-certificate, Repeat = students must repeat grade after obtaining a B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining a B-certificate. SES gap at graduation: difference in percentage college graduates between high and low SES. Opportunity cost of time: $10/h. Bootstrap standard errors in parentheses.

Table 6: Counterfactual tracking policies: differences with the pure discrete choice model

	Changes after repeat policy				Changes after downgrade policy			
Effort as a choice variable	Yes	No	Difference		Yes	No	Difference	
Educational outcomes (%):								
Final high school track								
Academic	0.17	-0.61	-0.78	(0.19)	-1.13	-1.88	-0.75	(0.18)
Middle-theoretical	-0.96	-1.32	-0.35	(0.19)	-1.52	-1.88	-0.36	(0.17)
Middle-practical track	-1.13	-1.73	-0.60	(0.17)	-0.77	-1.00	-0.23	(0.16)
Vocational track	-2.01	-1.04	0.96	(0.24)	5.03	6.21	1.19	(0.20)
Dropout	3.94	4.70	0.77	(0.20)	-1.61	-1.46	0.15	(0.12)
At least 1 B-certificate	-10.23	-6.32	3.91	(0.44)	-3.49	-0.86	2.63	(0.29)
At least 1 C-certificate	-0.23	0.22	0.44	(0.20)	-2.15	-1.69	0.45	(0.15)
At least 1 year of study delay	9.48	11.61	2.13	(0.36)	-9.82	-9.19	0.63	(0.27)
Higher education								
Enrollment	-1.76	-3.02	-1.27	(0.17)	-1.40	-2.27	-0.87	(0.11)
Graduation	-1.70	-2.69	-0.99	(0.15)	-0.30	-1.12	-0.81	(0.09)
SES gap at graduation	0.06	0.06	0.00	(0.21)	0.11	-0.15	0.26	(0.17)
Student welfare ($1000):								
Fixed costs (-)	0.85	1.26	0.41	(0.12)	-0.48	-0.63	-0.14	(0.05)
Variable costs (-)	0.49	0.00	-0.49	(0.08)	0.21	0.00	-0.21	(0.03)
Expected payoff (+)	-0.65	-0.95	-0.30	(0.06)	0.32	0.14	-0.17	(0.04)
Taste shocks (+)	-0.15	-0.02	0.12	(0.06)	-1.61	-1.76	-0.14	(0.07)
Total	-2.14	-2.23	-0.09	(0.09)	-1.02	-0.99	0.03	(0.05)

NOTE. – Predictions of two dynamic models. In a pure discrete choice model, students cannot adjust study effort. In the proposed model they can because they choose the distribution of performance through their choice of effort. Changes are with respect to the status quo prediction of each model. C-certificate: repeat grade. B-certificate = students acquired skills to proceed to next grade but only if they downgrade, i.e. switch to track of lower academic level or drop an elective course. Status quo = students can choose to downgrade or repeat grade after obtaining a B-certificate, Repeat = students must repeat grade after obtaining a B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining a B-certificate. SES gap at graduation: difference in percentage college graduates between high and low SES. Expected payoff = Expected payoff after high school. Opportunity cost of time: $10/h. Bootstrap standard errors in parentheses.

7 Conclusion

I estimated a dynamic model of effort choice in secondary education in which students choose the academic level of the study program, as well as the distribution of their performance. I find that policies that encourage underperforming students to choose programs of lower academic level do not reduce the number of higher education graduates. Moreover, they decrease grade retention and high school dropout which creates large savings for society that can be reinvested in early childhood education.

Future research could combine this approach with recent extensions of the pure discrete choice model along other dimensions, introducing additional uncertainty about the performance distribution due to imperfect information of students about their ability (Arcidiacono et al., 2023) and endogenous quality of schools or programs due to the quality of peers and effort choices of teachers (Fu and Mehta, 2018). This would allow for counterfactuals that change the educational system more substantially, such as deferring the age of tracking.

From a methodological perspective, I show that it is possible to allow students to exert different amounts of effort in counterfactual simulations by exploiting only commonly available data on program choices and performance outcomes. Further research can apply this strategy in other contexts where agents are expected to have some, but imperfect, control over state transitions. It would also be interesting to test the performance of both models by using exclusion restrictions (as discussed in Appendix A.6).

References

Abbring, Jaap H. 2010. Identification of dynamic discrete choice models. *Annual Review of Economics* 2 (1):367–394.

Aguirregabiria, Victor and Pedro Mira. 2010. Dynamic discrete choice structural models: A survey. *Journal of Econometrics* 156 (1):38–67.

Ahn, Thomas, Peter Arcidiacono, Amy Hopson, and James Thomas. 2022. Equilibrium grade inflation with implications for female interest in STEM majors. NBER Working Paper 26556, National Bureau of Economic Research, Cambridge, MA.

Altonji, Joseph G. 1995. The effects of high school curriculum on education and labor market outcomes. *The Journal of Human Resources* 30 (3):409–438.

Altonji, Joseph G., Erica Blom, and Costas Meghir. 2012. Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics* 4 (1):185–223.

Angrist, J. D. and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion.* Princeton: Princeton University Press.

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

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

Arcidiacono, Peter, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. 2023. College attrition and the dynamics of information revelation. NBER Working Paper 22325, National Bureau of Economic Research, Cambridge, MA.

Arcidiacono, Peter and Paul B. Ellickson. 2011. Practical methods for estimation of dynamic discrete choice models. *Annual Review of Economics* 3 (1):363–394.

Arcidiacono, Peter and Robert A. Miller. 2011. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* 79 (6):1823–1867.

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

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

Bhargava, Saurabh and George Loewenstein. 2015. Behavioral economics and public policy 102: Beyond nudging. *American Economic Review* 105 (5):396–401.

Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college. *International Economic Review* 44 (2):361–422.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil. 2011. Estimating marginal returns to education. *American Economic Review* 101 (6):2754–2781.

Cockx, Bart, Muriel Dejemeppe, Andrey Launov, and Bruno Van der Linden. 2018. Imperfect monitoring of job search: Structural estimation and policy design. *Journal of Labor Economics* 36 (1):75–120.

Cockx, Bart, Matteo Picchio, and Stijn Baert. 2019. Modeling the effects of grade retention in high school. *Journal of Applied Econometrics* 34 (3):403–424.

Costrell, Robert M. 1994. A simple model of educational standards. *American Economic Review* 84 (4):956–971.

Cummins, Joseph R. 2017. Heterogeneous treatment effects in the low track: Revisiting the Kenyan primary school experiment. *Economics of Education Review* 56:40–51.

Cunha, Flavio and James J. Heckman. 2009. The economics and psychology of inequality and human development. *Journal of the European Economic Association* 7 (2/3):320–364.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78 (3):883–931.

De Groote, Olivier and Koen Declercq. 2021. Tracking and specialization of high schools: Heterogeneous effects of school choice. *Journal of Applied Econometrics* 36 (7):898–916.

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

———. 2018. Enrollment and degree completion in higher education without admission standards. *Economics of Education Review* 66:223–244.

Dubois, Pierre, Alain de Janvry, and Elisabeth Sadoulet. 2012. Effects on school enrollment and performance of a conditional cash transfer program in Mexico. *Journal of Labor Economics* 30 (3):555–589.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review* 101 (5):1739–1774.

Dustmann, Christian, Patrick A. Puhani, and Uta Schoenberg. 2017. The long-term effects of early track choice. *The Economic Journal* 127 (603):1348–1380.

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

French, Eric and Christopher Taber. 2011. Identification of models of the labor market. In *Handbook of Labor Economics*, vol. 4. Elsevier, 537–617.

Freyberger, Joachim. 2018. Non-parametric panel data models with interactive fixed effects. *The Review of Economic Studies* 85 (3):1824–1851.

Fruehwirth, Jane Cooley, Salvador Navarro, and Yuya Takahashi. 2016. How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. *Journal of Labor Economics* 34 (4):979–1021.

Fu, Chao and Nirav Mehta. 2018. Ability tracking, school and parental effort, and student achievement: A structural model and estimation. *Journal of Labor Economics* 36 (4):923–979.

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

Gary-Bobo, Robert J., Marion Goussé, and Jean-Marc Robin. 2016. Grade retention and unobserved heterogeneity. *Quantitative Economics* 7 (3):781–820.

Gayle, George-Levi, Limor Golan, and Robert A. Miller. 2015. Promotion, turnover, and compensation in the executive labor market. *Econometrica* 83 (6):2293–2369.

Gayle, George-Levi and Robert A. Miller. 2015. Identifying and testing models of managerial compensation. *The Review of Economic Studies* 82 (3):1074–1118.

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

Goodman, Joshua. 2019. The labor of division: Returns to compulsory high school math coursework. *Journal of Labor Economics* 37 (4):1141–1182.

Guyon, Nina, Eric Maurin, and Sandra McNally. 2012. The effect of tracking students by ability into different schools: A natural experiment. *Journal of Human Resources* 47 (3):684–721.

Hanushek, Eric A. and Ludger Woessmann. 2006. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal* 116 (510):C63–C76.

Heckman, James J., John Eric Humphries, and Gregory Veramendi. 2016. Dynamic treatment effects. *Journal of Econometrics* 191 (2):276–292.

Heckman, James J. and Stefano Mosso. 2014. The economics of human development and social mobility. *Annual Review of Economics* 6 (1):689–733.

Heckman, James J. and Salvador Navarro. 2007. Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics* 136 (2):341–396.

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

Hu, Yingyao and Matthew Shum. 2012. Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics* 171 (1):32–44.

Hu, Yingyao and Yi Xin. 2022. Identification and estimation of dynamic structural models with unobserved choices. Cemmap working paper, No. CWP44/20, The Institute for Fiscal Studies, Department of Economics, UCL.

Jacob, Brian A and Lars Lefgren. 2009. The effect of grade retention on high school completion. *American Economic Journal: Applied Economics* 1 (3):33–58.

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

Joensen, Juanna Schrrter and Elena Mattana. 2021. Student aid, academic achievement, and labor market behavior: Grants or Loans? Working paper.

Kalouptsidi, Myrto, Paul T. Scott, and Eduardo Souza-Rodrigues. 2021. Identification of counterfactuals in dynamic discrete choice models. *Quantitative Economics* 12 (2):351–403.

Kapor, Adam J., Christopher A. Neilson, and Seth D. Zimmerman. 2020. Heterogeneous beliefs and school choice mechanisms. *American Economic Review* 110 (5):1274–1315.

Kasahara, Hiroyuki and Katsumi Shimotsu. 2009. Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* 77 (1):135–175.

Keane, Michael P. and Kenneth I. Wolpin. 1997. The career decisions of young men. *Journal of Political Economy* 105 (3):473–522.

Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen. 2015. Behavioral economics of education. *Journal of Economic Behavior & Organization* 115:3–17.

Kreisman, Daniel and Kevin Stange. 2020. Vocational and career tech education in American high schools: The value of depth over breadth. *Education Finance and Policy* 15 (1):11–44.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach. 2018. School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics* 10 (2):1–26.

Lavecchia, Adam, Heidi Liu, and Philip Oreopoulos. 2016. Behavioral economics of education: Progress and possibilities. In *Handbook of the Economics of Education*, vol. 5. National Bureau of Economic Research, 1–74.

Lin, Dajun. 2020. Multi-dimensional abilities, task content of occupations and career choices: A dynamic analysis. Working paper.

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

Manacorda, Marco. 2012. The cost of grade retention. *Review of Economics and Statistics* 94 (2):596–606.

Margiotta, Mary M. and Robert A. Miller. 2000. Managerial compensation and the cost of moral hazard. *International Economic Review* 41 (3):669–719.

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

Nord, C, S Roey, R Perkins, M Lyons, N Lemanski, J Brown, and J Schuknecht. 2011. The nation's report card: America's high school graduates. Tech. Rep. NCES 2011-462, US Department of Education, National Center for Education Statistics, Washington, DC.

OECD. 2012. *Education at a glance 2012*. OECD Publishing.

OECD. 2013. *PISA 2012 results: What makes schools successful*, vol. IV. OECD Publishing.

———. 2017. *Education at a glance 2017*. OECD Publishing.

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

Pekkarinen, Tuomas, Roope Uusitalo, and Sari Kerr. 2009. School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics* 93:965–973.

Roller, Marcus and Daniel Steinberg. 2020. The distributional effects of early school stratification - non-parametric evidence from Germany. *European Economic Review* 125.

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

Rust, John. 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica* 55 (5):999–1033.

Schneeweis, Nicole and Martina Zweimüller. 2014. Early tracking and the misfortune of being young. *The Scandinavian Journal of Economics* 116 (2):394–428.

Shure, Nikki. 2021. Non-cognitive peer effects in secondary education. *Labour Economics* 73.

Todd, Petra E and Kenneth I Wolpin. 2018. Accounting for mathematics performance of high school students in Mexico: Estimating a coordination game in the classroom. *Journal of Political Economy* 126 (6):2608–2650.

Train, Kenneth. 2009. *Discrete choice methods with simulation*. Cambridge; New York: Cambridge University Press.

van den Berg, Gerard J. and Bas van der Klaauw. 2019. Structural empirical evaluation of job search monitoring. *International Economic Review* 60 (2):879–903.

A Identification

A.1 Identified objects in pure discrete choice models

Consider the standard setup of a dynamic discrete choice model. Each period t, agent i chooses an option j. The decision is based on observed characteristics x_{it}, an unobserved type ν_i and unobserved iid taste shocks $\varepsilon_{it} = \{\varepsilon_{i1t}, \varepsilon_{i2t}, ...\}$. The time horizon can be infinite or (if x_{it} includes t) finite. Each period agent i derives some flow utility

$$u_j(x_{it}, \nu_i) + \varepsilon_{ijt}$$

and states transition according to a process that satisfies conditional independence (Rust, 1987):

$$f_j(x_{it+1}|x_{it}, \nu_i) = f_j(x_{it+1}|x_{it}, \nu_i, \varepsilon_{ijt}).$$

Agents maximize expected lifetime utility by choosing the option with the highest conditional value function:

$$v_j(x_{it}, \nu_i) + \varepsilon_{ijt} = u_j(x_{it}, \nu_i) + \beta \int \bar{V}(x_{it+1}, \nu_i) f_j(x_{it+1}|x_{it}, \nu_i) dx_{it+1} + \varepsilon_{ijt}$$

with $\bar{V}(x_{it+1}, \nu_i)$ the expected value of behaving optimally after integrating over the taste shocks.

In the case where there is no unobserved type, Magnac and Thesmar (2002) show that data on x_{it} and the chosen option, identify $u_j(x_{it})$ after specifying the utility of a reference alternative, the discount factor β and the distribution of ε_{ijt}. State transitions $f_j(x_{it+1}|x_{it})$ are nonparametrically identified. We could then use $u_j(.)$ and $f_j(.)$ for counterfactual simulations by assuming they are invariant to policy changes.

As the iid assumption on unobserved heterogeneity is restrictive, many applications would add an unobserved state ν_i to capture persistent unobserved heterogeneity and identify $u_j(x_{it}, \nu_i)$ and $f_j(x_{it+1}|x_{it}, \nu_i)$. I will allow for this in the rest of this section.[24]

A.2 Identification of policy-invariant functions

To relax the assumption of policy-invariance, assume instead that the functions $u_j(.)$ and $f_j(.)$ depend on choice behavior and what we identify are therefore the endogenously determined objects $u_j^*(x_{it}, \nu_i)$ and $f_j^*(x_{it+1}|x_{it}, \nu_i)$. The goal is to derive other transformations of the data that are more likely to be policy-invariant.

Assume agents can choose the distribution of state transitions through a single index y_{it} such that $\phi_{j,\tilde{x},\widetilde{x}'}(y_{it})$ is the probability for i in state $(x_{it}, \nu_i) = \tilde{x}$ to transition to state $(x_{it+1}, \nu_i) = \widetilde{x}'$ after choosing j. The optimal choice of y_{it} in a given program j and state (x_{it}, ν_i) is then given by $y_j^*(x_{it}, \nu_i) = \phi_{j,\tilde{x},\widetilde{x}'}^{-1}(f_j^*(x_{it+1}|x_{it}, \nu_i))$.[25] We now let this index linearly enter the utility function:

$$u_j(x_{it}, \nu_i, y_{it}) = u_j^0(x_{it}, \nu_i) + u_j^y(x_{it}, \nu_i)y_{it} \tag{16}$$

with $u_j^0(x_{it}, \nu_i)$ a component that is independent of the choice of the index and $u_j^y(x_{it}, \nu_i) \equiv \frac{du_j(x_{it},\nu_i,y_{it})}{dy_{it}}$ the marginal flow utility from changing y_{it}.

To connect what we observe in the data with the current model, we make the following assumption:

Assumption: In the data, agents in option j choose y_{it} to maximize expected lifetime utility and obtain an interior solution $y_j^*(x_{it}, \nu_i) = \phi_{j,\tilde{x},\widetilde{x}'}^{-1}(f_j^*(x_{it+1}|x_{it}, \nu_i))$ with $\phi_{j,\tilde{x},\widetilde{x}'}$ a known function that is invertible and differentiable in y_{it}.

[24] There are several approaches to identify type-specific $u_j(x_{it}, \nu_i)$ and $f_j(x_{it+1}|x_{it}, \nu_i)$, see for example Magnac and Thesmar (2002); Kasahara and Shimotsu (2009); Hu and Shum (2012), and the discussion on identification of the application in this paper.

[25] For a simple case, assume $x_{it} = (x_{i0}, g_{it})$ with initial observed characteristic x_{i0} and a dummy for obtaining a degree g_{it}. Assume students choose the exponential of the index of a logit on obtaining a degree when they are in an option j that gives this possibility. We can then write $\phi_{j,(x_{i0},0),\nu_i,(x_{i0},1)}(\tilde{y}) = \frac{\tilde{y}}{1+\tilde{y}}$ and $\phi_{j,(x_{i0},0),\nu_i,(x_{i0},0)}(\tilde{y}) = 1 - \frac{\tilde{y}}{1+\tilde{y}}$.

In contrast, the pure discrete choice model recovers $f_j^*(x_{it+1}|x_{it},\nu_i)$ while remaining agnostic about how it was determined. However, when proceeding to counterfactual simulations, it is not updated, i.e. it is implicitly assumed that agents cannot affect it.

As in the pure discrete choice model, we still assume that agents choose the option j that generates the highest expected lifetime utility. Conditional value functions now depend on y_{it}:

$$v_j(x_{it},\nu_i,y_{it})+\varepsilon_{ijt}=u_j(x_{it},\nu_i,y_{it})+\beta\int\bar{V}(x_{it+1},\nu_i)\phi_{j,\widetilde{x},\widetilde{x}'}(y_{it})dx_{it+1}+\varepsilon_{ijt}.$$

Solving this for the optimal y_{it}, the following FOC has to be satisfied:

$$u_j^y(x_{it},\nu_i)=-\beta\int\bar{V}(x_{it+1},\nu_i)\frac{\partial\phi_{j,\widetilde{x},\widetilde{x}'}(y_{it})}{\partial y_{it}}dx_{it+1}\text{ for }y_{it}=y_j^*(x_{it},\nu_i)\qquad(17)$$

with the left-hand side equal to the marginal flow utility i receives today from increasing y_{it}, and the right-hand side the expected decrease in future utility. Since $\phi_{j,\widetilde{x},\widetilde{x}'}$ is invertible and differentiable, we can identify the optimal value of y_{it} in the data ($y_j^*(x_{it},\nu_i)$) and calculate the derivative at this point. Identification of all flow utilities with optimal choices also implies the identification of $\bar{V}(x_{it+1},\nu_i)$. β is taken as given. Therefore, we can identify $u_j^y(x_{it},\nu_i)$ using this FOC. With $u_j^y(x_{it},\nu_i)$, $y_j^*(x_{it},\nu_i)$ and $u_j^*(x_{it},\nu_i)$ identified, we can use (16) to identify $u_j^0(x_{it},\nu_i)$.

We have now identified two new functions of state variables in the model: the marginal utility of a change in the index of state transitions $u_j^y(x_{it},\nu_i)$, and a component in the utility function that is independent of the distribution of state transitions $u_j^0(x_{it},\nu_i)$. We can do this for different choices of $\phi_{j,\widetilde{x},\widetilde{x}'}(y_{it})$, giving some flexibility to researchers to choose what remains fixed in counterfactual simulations. Two aspects are important for this choice. First, by choosing $\phi_{j,\widetilde{x},\widetilde{x}'}$, the researcher effectively chooses for which transformation of state transitions the linearity assumption holds, i.e. for which transformation the marginal impact on flow utility can be considered policy-invariant. Second, the choice of $\phi_{j,\widetilde{x},\widetilde{x}'}$ should be consistent with a high-level assumption that the FOC (17) is satisfied in the data (as an interior solution is required). If not, the FOC cannot provide the identifying power we need. Note that we only need this for identification. In counterfactual simulations, we can allow for corner solutions.

A.3 Extension: a vector of choice variables

With multiple stochastic variables, $\phi_{j,\widetilde{x},\widetilde{x}'}(y_{it})$ is a joint probability that depends on a scalar y_{it}. Alternatively, we could also invert each of them separately and obtain a vector of choice variables $y_{it}' = (y_{it}^1, y_{it}^2, ..., y_{it}^{G_y})$, with G_y, the number of stochastic variables of which the distribution is chosen by i. We can extend the linearity assumption in the utility function to each element of y_{it}. This yields the same equation (16) but with y_{it} a vector. We can then identify a vector for the parameters $u_j^y(x_{it}, \nu_i) = (u_j^{y^1}(x_{it}, \nu_i), u_j^{y^2}(x_{it}, \nu_i), ...)$ by using the following first-order system of equations:

$$u_j^{y^1}(x_{it}, \nu_i) = -\beta \int \bar{V}(x_{it+1}, \nu_i) \frac{\partial \phi_{j,\widetilde{x},\widetilde{x}'}(y_{it})}{\partial y_{it}^1} dx_{it+1} \tag{18}$$

$$u_j^{y^2}(x_{it}, \nu_i) = -\beta \int \bar{V}(x_{it+1}, \nu_i) \frac{\partial \phi_{j,\widetilde{x},\widetilde{x}'}(y_{it})}{\partial y_{it}^2} dx_{it+1}$$

$$... \tag{19}$$

$$\text{for } y_{it} = y_j^*(x_{it}, \nu_i) \tag{20}$$

Note that in most cases (also when y_{it} is a scalar) it will not be possible to obtain closed-form solutions for y_{it}^*. While the status quo optimum can be recovered from the data $(y_j^*(x_{it}, \nu_i))$, counterfactual simulations will still require computationally-intensive methods (e.g. a grid search) which becomes more intensive when y_{it} is a vector. This is why in the application we use a singleton y_{it} by integrating it out of the distribution of track restrictions. We then estimate how other performance outcomes depend on it and we allow for a (policy-invariant) comparative advantage.

A.4 Extension: long-lasting effects of y_{it}

In certain cases we can allow for a direct impact of y_{it} beyond $t+1$. This is for example useful when performance only proxies the accumulated human capital and misses an investment that would only realize in future outcomes. Consider a utility function which depends on the past level of y_{it} in a flexible way (a similar approach can be applied to allow for further lags, or to take into account some (weighted) average of past levels of y_{it}):

$$u_j(x_{it}, \nu_i, y_{it}, y_{it-1}) = u_j^0(x_{it}, \nu_i, y_{it-1}) + u_j^y(x_{it}, \nu_i, y_{it-1})y_{it} \tag{21}$$

This leads to the following conditional value function:

$$v_j(x_{it}, \nu_i, y_{it}, y_{it-1}) + \varepsilon_{ijt} = u_j(x_{it}, \nu_i, y_{it}, y_{it-1}) + \beta \int \bar{V}(x_{it+1}, \nu_i, y_{it})\phi_{j,\tilde{x},\tilde{x}'}(y_{it})dx_{it+1} + \varepsilon_{ijt}.$$

In period t, i solves the following FOC, given the optimal value that was chosen in $t-1$:

$$u_j^y(x_{it}, \nu_i, y_{it-1}^*) = -\beta \int \left(\frac{\partial \bar{V}(x_{it+1}, \nu_i, y_{it})}{\partial y_{it}} \phi_{j,\tilde{x},\tilde{x}'}(y_{it}) + \bar{V}(x_{it+1}, \nu_i, y_{it}) \frac{\partial \phi_{j,\tilde{x},\tilde{x}'}(y_{it})}{\partial y_{it}} \right) dx_{it+1}$$
$$\text{for } y_{it} = y_j^*(x_{it}, \nu_i, y_{it-1}^*) \ .$$

A first difference is that we now need to identify $y_{it}^* = y_j^*(x_{it}, \nu_i, y_{it-1}^*)$, rather than $y_{it}^* = y_j^*(x_{it}, \nu_i)$ from the data. Assuming we observe the first period in which effort can be exerted, we can first identify $y_{j1}^* = y_j^*(x_{i1}, \nu_i)$ and use it as a predictor for $y_{j2}^* = y_j^*(x_{i2}, \nu_i, y_{i1}^*)$. Note that y_{j2}^* will only be different if y_{i1}^* depends on other variables than (x_{i2}, ν_i), i.e. if $x_{i1} \neq x_{i2}$. However, we only need to know the predicted value y^*_{j2}, implying that we do not need time-varying state variables here as we do not need to distinguish the effect of y^*_{i1} from the effect of x_{i2}.

A second difference is that we need to calculate the derivative of the ex-ante value function with respect to y_{it}: $\frac{\partial \bar{V}(x_{it+1},\nu_i,y_{it})}{\partial y_{it}}$. This follows from the fact that future values, and not only state transitions, are influenced by the current level of y_{it}. Here we do need to distinguish between the effect of y_{it} and x_{it+1}, and therefore require $x_{it} \neq x_{it+1}$. In certain applications, this derivative can be easy to find. In the example of section 2, we used extreme value type 1 error terms which result in a simple closed-form expression for the ex-ante value function. Moreover, we explained that if finite dependence (one period ahead) holds, we can write it is as a function of the CCPs only. We now need the CCPs to be predicted by $(x_{it+1}, \nu_i, y^*_{it})$, and differences arise when $x_{it} \neq x_{it+1}$. Moreover, we need to know $\frac{\partial \ln \Pr(d_{it+1}=0|x_{it+1},\nu_i,y_{it})}{\partial y_{it}}|_{y_{it}=y^*_{it}}$, which illustrates the need to separately identify the effect of y^*_{it} from x_{it+1}.

Similar as before, we identify $u^0_j(x_{it}, \nu_i, y^*_{it-1})$ by rewriting (21) at the optimal level y^*_{it}. We identified all y^*_{it} and $u^y_j(x_{it}, \nu_i, y^*_{it-1})$ in the previous paragraph. $u_j(x_{it}, \nu_i, y^*_{it}, y^*_{it-1})$ is what is identified using the standard Magnac and Thesmar (2002) conditions, provided we now also include the state variables that predict y^*_{it-1}.

Note that identification and estimation can be achieved here without having to solve the full model, however, this is needed for counterfactuals. This is made substantially more complicated by introducing a dependence on past values of y_{it}. If we solve the model by backward induction, we need to know the optimal choices in $t+1$, conditional on each value of y_{it}, instead of only keeping track of the state it changes in $t+1$ (in the application: the choice set restrictions or the high school degree). This is the main reason to not pursue this in the application of the paper. A concern there is that higher education outcomes might be directly affected by the effort exerted all through high school, while we only capture it in the way it influences their high school programs and the years of study delay. Appendix F shows that this is likely not the case here. To do this, I exploit additional test score data to show that past values of effort are unlikely to have an important impact on higher education outcomes, beyond what we do allow for in the model.

A.5 Example with continuous outcome

The toy model in section 2 shows an example of a two-period model with a binary outcome, while the application allows for a vector of discrete outcomes and multiple periods and alternatives. To clarify the generality of the approach, this section discusses the case of a continuous outcome variable. This specification is similar to Ahn et al. (2022), but without using a proxy for effort. Moreover, the model in Ahn et al. (2022) is static, while I use the dynamics here to identify the benefit of a higher grade.

In this example I make two changes to the model of section 2: first, assume the outcome variable is a continuous test score g_i. We normalize the test score by subtracting the mean and dividing it by its standard deviation. From the perspective of a student, we assume this score follows a normal distribution with mean $\ln(y_i)$ and standard deviation σ: $N(\ln(y_i), \sigma^2)$. $y_i \in (0, +\infty)$ is the effective study effort and allows students to set the expected value of grades. σ is a parameter to be estimated, assumed to be policy-invariant. As the optimal value is the same for students with the same state variable in the initial period we can write $y_i^* = y^*(x_i)$. σ captures the predictability of test scores.

Second, as g_i no longer captures degree completion, we assume anyone can enroll in higher education after attending high school, removing the dependence of the choice set on g_i. Instead, we allow the utility of enrolling in college to be a function of grades: $\Psi_1(x_i, g_i)$, with $\frac{\partial \Psi_1(x_i,g_i)}{\partial g_i} > 0$. This could capture a higher chance to be accepted in college, or a higher utility when better prepared. It can be estimated using choice data for students that obtained different grades.

y_i is chosen optimally in the conditional value function of $j = 1$:

$$v(x_i, y_i) = -C^0(x_i) - c(x_i)y_i + \beta\gamma + \beta \int \left(\ln\left(1 + \exp \Psi_1(x_i, g_i)\right)\right) \phi_g(\ln(y_i), \sigma^2) dg \quad (22)$$

with $\phi_g(\ln(y_i), \sigma^2)$ the density of the grade distribution. As in the discrete context, I specified the distribution of performance to capture that it is generally harder to improve an outcome if it is already high. I assume a constant marginal cost for a unit increase in y_i, which is the exponential transformation of the expected grade. At the mean of the overall grade distribution, $c(x_i)$ is the marginal cost to improve the grade. However, at higher values, paying $c(x_i)$ leads to smaller increases in the grade.[26]

A solution for y_i has to satisfy the following FOC:

$$c(x_i) = \beta \int \left(\ln\left(1 + \exp \Psi_1(x_i, g_i)\right)\right) \frac{\partial \phi_g(\ln(y_i), \sigma^2)}{\partial \ln(y_i)} \frac{1}{y_i} dg \text{ if } y_i{=}y_i^*$$

The right-hand side captures the marginal benefits and can be found using the density function of a normal distribution: $\frac{\partial \phi_g(\ln y_i, \sigma^2)}{\partial \ln y_i} = \frac{(g - \ln y_i)\exp(-0.5(\ln y_i - g)^2/\sigma^2)}{\sigma^3\sqrt{2\pi}}$. As in the discrete case, ruling out $y_i = 0$, implies an upper bound on the marginal cost estimates. We also need to rule out infinitely large y_i. Costs grow exponentially in the expected grade. Also, when y_i grows, $\frac{\partial \Psi_1(x_i, g_i)}{\partial g_i} > 0$ ensures that $\ln\left(1 + \exp \Psi_1(x_i, g_i)\right) \to \Psi_1(x_i, g_i)$. A sufficient condition to rule out infinite y_i is therefore to impose weakly decreasing returns ($\frac{\partial^2 \Psi_1(x_i, g_i)}{\partial g_i^2} \leq 0$), e.g. a constant return to g_i.[27]

While continuous state variables are often used in dynamic discrete choice models,

[26] Note that $\frac{\partial g_i}{\partial y_i} = \frac{1}{y_i}$ and the mean of the grade distribution (0) is obtained by setting $y_i = 1$.

[27] Alternatively, we can choose a truncated normal distribution for the grades, with a truncation by the minimum and maximum possible test score. In this case, y_i is more difficult to interpret as $\ln(y_i)$ is no longer the mean of that distribution. However, it is still the mode if it does not cross the boundaries of the distribution. A benefit from this distribution is that we do not need to assume decreasing returns to test scores. As in the discrete case, benefits are bounded by the maximum grade, while y_{it} remains unbounded, meaning that marginal benefits and marginal costs will intersect as long as $\frac{\partial \Psi_1(x_i, g_i)}{\partial g_i} > 0$.

they usually require discretization or interpolation to be solved (Aguirregabiria and Mira, 2010). Moreover, identification results make use of discrete state variables (Magnac and Thesmar, 2002). It is therefore useful to write the model by discretizing the grade distribution in K bins. The conditional value functions are then:

$$\begin{aligned} v(x_i, y_i) &= -C^0(x_i) - c(x_i)y_i + \beta\gamma \\ &\quad + \beta \sum_{k=1}^{K} \left(\left(F(g_k) - F(g_{k-1})\right) \left(\ln\left(1 + \exp \Psi_1(x_i, g_i = g_k)\right)\right)\right) \end{aligned}$$

with $F(g_k)$ the cdf of the grade distribution evaluated at g_k, with $F(g_0) = 0$ and $F(g_K) = 1$. Similarly, the FOC can be derived:

$$\begin{aligned} c(x_i) &= \beta \frac{1}{y_i} \sum_{k=1}^{K} \left(\frac{dF(g_k)}{d\ln y_i} - \frac{dF(g_{k-1})}{d\ln y_i} \right) \left(\ln\left(1 + \exp \Psi_1(x_i, g_i = g_k)\right)\right) \\ &\text{if } y_i{=}y_i^*. \end{aligned}$$

with $\frac{dF(g_k)}{d\ln y_i} = -\frac{1}{\sigma}\tilde{\phi}(\frac{g_k - \ln y_i}{\sigma})$ and $\tilde{\phi}$ the pdf of a standard normal distribution.

Note that the optimal level of effort y_i^* is identical for students in the same state: $y_i^* = y^*(x_i)$. With grades distributed $N(\ln(y_i), \sigma^{\mathbf{2}})$, $y^*(x_i)$ can be obtained from an OLS regression of grades on a flexible function of x_i. The estimated mean is then $\ln y^*(x_i)$, while the variance of the error term provides an estimate of σ^2. In a counterfactual, we keep σ^2 fixed but look for new values of y_i^*.

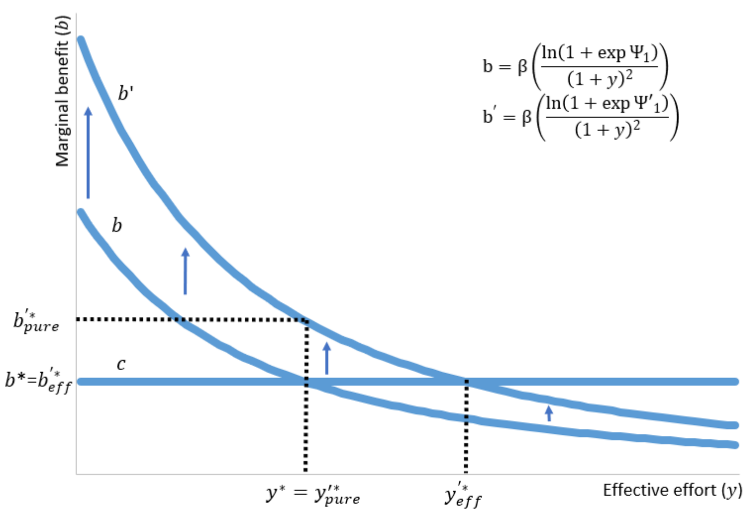
Finally, an alternative model can be proposed that implies discretization from the very start. The ordered logit structure that is used in the application in this paper can also be used for grades that take many values. The main difference is the way the linearity assumption in the utility function is implemented. The current model implies a constant marginal cost in the exponential transformation of the expected value of grades. The ordered logit structure instead assumes a constant marginal cost to increase the odds of avoiding the lowest outcome.

A.6 The difference with a pure discrete choice model

A pure discrete choice model (in the simple context discussed in section 2) will assume that $u^*(x_i)$ and $y^*(x_i)$ are policy-invariant functions. In this paper we treat them as endogenous outcomes, with two alternative model components assumed to be policy-invariant: fixed costs $C^0(x_i)$ and marginal costs $c(x_i)$.

The importance of this change is easy to see when we consider a counterfactual simulation that raises the value of college ($\Psi_1(x_i)$). In both models, this would have a direct impact on college enrollment among high school graduates through (2). In a dynamic model, it also impacts the school choice in $t = 1$ because students know a high school degree is required to enter college (see (3)). However, only the effort model allows the probability to obtain a degree $\phi(y_i)$ to be affected too. Figure A1 shows the difference between models graphically for a given value of x_i, which I omit in the discussion. The marginal benefit of increasing y is $b = \beta\left(\frac{\ln(1+\exp\Psi_1)}{(1+y)^2}\right)$.

Figure A1: Equilibrium effort in status quo and counterfactual



Note: Marginal costs (c) and benefits (b) of different levels of effort (y) for a given value of individual characteristics. The upward shift in marginal benefits simulates a counterfactual increase in the value of better performance through the value of college. Stars denote optimal values. Pure denotes the counterfactual optimum in the pure discrete choice model, eff denotes the counterfactual optimum in the effort model.

We can identify y^* from the data, which chooses a point on the marginal benefit curve: b^*. Now consider a counterfactual increase in the college value: $\Psi_1' > \Psi_1$. This shifts the marginal benefit curve upwards: $b' > b$. The effort model assumes a policy-invariant marginal cost that should equal the marginal benefits: $b_{eff}'^* = b^*$. Therefore, effort increases: $y_{eff}'^* > y^*$. Instead, a pure discrete choice models assumes a policy-invariant effort level: $y_{pure}'^* = y^*$. Note that this implies $b_{pure}'^* > b^*$. If there is such policy variation in the data, we could test which of the two sets of conditions holds. Similarly, exclusion restrictions could be used. Let z_i be a variable affecting future values (here the college value), but not marginal costs. This will cause a shift of the marginal benefits curve. The effect of z_i could be non-monotonic, especially in more complex models, thereby making it difficult to learn from inequalities. Still, a sufficient condition to reject the pure discrete choice model would be $y^*(x_i, z_i) \neq y^*(x_i)$. This test is straightforward because obtaining estimates of $y^*(x_i, z_i)$ is a first step in estimating either model. Therefore, we can look at the statistical significance of z_i. To reject the effort model, a sufficient condition would be $b^*(x_i, z_i) \neq b^*(x_i)$.

Alternatively, the analysis shows we need to choose between two sets of assumptions: a pure discrete choice model that assumes performance is exogenous to policy changes, or a dynamic model of effort decisions that assumes students optimally choose the probability to perform well at a marginal cost that is policy-invariant.[28] The former is best suited in a context where performance outcomes mainly serve to measure a level of ability or knowledge. For high-stakes exams, assuming students optimize the probability to do well is more reasonable. We can also take a conservative stance and consider that both models put reasonable bounds on counterfactual predictions. As can be seen in Figure A1, the pure discrete choice model can also be

[28] As in Keane and Wolpin (1997), one could also avoid modeling performance and ignore a year of schooling that is not successfully completed, i.e. $d_{i1} = 1$ only if we also observe $g_i = 1$. This way, we do not need to take stance on the process of g_i. A counterfactual change in d_{i1} now also captures the students that were already in school but needed this extra incentive to get a degree. The problem with this approach is that we do not observe how the probability to obtain a degree changes because of the policy and we cannot run counterfactuals that change the implications of that.

interpreted as an effort model with a perfectly inelastic marginal cost curve, instead of a perfectly elastic one. A marginal cost curve that increases with the level of effort would result in a counterfactual level in between the two equilibria. However, since effort was defined as the odds of success, it is unclear if we should expect increasing marginal costs as this choice of functional form was already motivated by the fact that increasing the probability of success is likely harder when it is already high.

Because there is no policy variation in the data, I cannot test that the model of effort choice performs better in this dataset. However, there is a variable that is reasonable to exclude from marginal costs and could shift effort: distance to college. As explained in section 4.8, I estimate levels of effort in the data as the exponential transformation of the index of an ordered logit model. The result of this ordered logit model is given in Table J10. Several measures of distance to higher education are indeed statistically significant. However, I do not make strong claims in the context of these data as distance to college is not a strong predictor of effort because all students live close to a college. I also cannot exclude they affect the marginal benefits in the data. Note that this is difficult here as the marginal benefits are the estimated marginal costs of the effort model, which are recovered from (15). This depends on flexible, but parametric, approximations of both CCPs and effort which both depend on distance to college. It is therefore unlikely the approximation is sufficiently good to exactly cancel out the effect on effort by the effect on the CCPs.

In other contexts, it has been shown that performance responds to policy changes because of changes in study effort (Costrell, 1994; Dubois et al., 2012; Garibaldi et al., 2012). Consider for example Dubois et al. (2012) who study the impact of a cash transfer experiment on a dummy performance outcome, required to transfer to the next grade. For the treatment group, transfers increased with school grade and stopped after graduation. The authors show in a theoretical model of effort choice that this created a dynamic incentive to perform well in early grades, as the increase in value creates incentives similar to our example of increasing college value. In later grades there is an opposite effect as the transfer stops after graduation, giving an incentive to repeat grades to stay in school. The theoretical results are confirmed in the experimental data.

B Data appendix

B.1 The LOSO dataset

The dataset used for this paper is the LOSO dataset.[29] The first part of the data contains rich information about students and their parents, and choices and performance measures during high school in the region of Flanders (Belgium). We can follow a cohort of students starting high school in 1990. I also include results from follow-up research, called “LOSO-annex”, which looked into the education and labor market career in the first three years after leaving high school (academic years starting in 1996 until 1998 for most students, but later for those with study delay). This data was later enriched by sending questionnaires during 2003-2005 to students that were still in the educational system in the questionnaire before.

The students are not randomly selected over Flanders. Instead, two large subregions of Flanders were defined that are considered to be representative of the entire region.[30] In these regions, almost all schools are included, and within each school, every student is included. The first subregion is in the east part of Flanders and includes the municipalities Hasselt, Genk, Beringen, Leopoldsburg, Herk-de-Stad, and Diest. The second subregion is more to the west and contains the schools in Dendermonde, Hamme, and Zele. Data was collected from students, parents, teachers, and schools, and they were actively contacted by researchers on multiple occasions. This is why the data is of high quality and there is very little attrition. Even if a student decides to leave his school for a school that was not initially part of the project, it was still possible to collect the necessary information.

[29] See also https://ppw.kuleuven.be/onderwijskunde/projecten/longitudinaal-onderzoek-schoolloopbananonderzoek/losodatabank.

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

B.2 Sample selection

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

The model in this paper captures the main aspects of the education system but also makes some simplifications, implying that it cannot explain every observation in the data. Moreover, some data on the choices or outcomes that are needed for the estimation are missing. Table B1 summarizes the attrition. More details on why observations had to be dropped follow next.

B.3 Data interpretation

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

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

Table B1: Data attrition

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

Certificates that exclude an entire track are straightforward to implement. This already contains 67% of the data on B-certificates. In other cases, I proceed as follows. I always assume a hierarchy: if a low track is excluded, the higher ones are excluded too.[31] I also use a slightly different definition of a B-certificate that is more consistent over the different grades. I ignore the officially called "C-certificates" in grade 7 as they do not restrict entry into grade 8 of the vocational track, and change them to B-certificates that allow the vocational track in the next grade (or A-certificate if the student is already in the vocational track). In other cases in the academic and middle-theoretical track, I use the following procedure. This procedure was established to be in line as much as possible with the spirit of the educational system, as well as to minimize the number of choices in the data that would not be possible to be explained by the model. I make groups of aggregated study programs that are less aggregated than the ones used in the model, but more aggregated than how they appear in the data. This aggregates over very small differences within programs between which a B-certificate is not expected to ever make a distinction, except when teachers (and probably students) are not aware of the existence of the program. A B-certificate then excludes all classical language options if all the aggregated programs with classical languages appear in the list of restrictions. It excludes math options and the entire track if there is an exclusion within all the major aggregated options of these study programs. For exclusion of the middle-practical track, one occurrence of a program in the track in the list of restrictions restricts the entire track, unless choice behavior and the corresponding grade are not consistent with that.

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

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

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

B.4 Details about study programs

The official distinction between tracks differs slightly from the one proposed in the paper. The official track names are "ASO", "TSO", "KSO", "BSO", and "BUSO" and the distinction for most tracks is made from the third year on (i.e. grade 9). ASO corresponds to the academic track, BSO and BUSO to the vocational track and both TSO and KSO are middle tracks (that differ in their focus on respectively technical education and artistic education). I then split up this middle track according to programs that prepare primarily for higher education (middle-theoretical) and the labor market (middle-practical), which is a common distinction made, e.g. in Cockx et al. (2019), but also by the researchers that collected the data.[32]

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

As mentioned in Cockx et al. (2019), upward mobility is theoretically possible but practically infeasible which is why it rarely occurs in the data. Nevertheless, I do allow for this flexibility in non-vocational tracks in the first two grades as I do

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

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

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

B.5 Distance and travel time data

I use the address data of students and schools to obtain coordinates using the Stata command "geocode3". For the schools, I updated this manually when geocode returned an error or was not very precise. I did this for schools with at least 10 student-time observations using Google maps. I then use the "osrmtime" command to calculate travel time by bike to the closest school that offers the study program.[33] Note that all schools attended by students in the sample are used, which includes also schools outside of the ones assigned by the researchers (because students can switch to other schools). I dropped students living more than 50km from any school as they are more likely to be influenced by schools that I do not observe or are outliers because of measurement error when geocoding.

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

B.6 Policy relevance

Although similar issues arise in other educational systems, they are particularly important in the current context. Belgium spends 2.8% of its GDP on secondary education, the highest number among OECD countries. Therefore, it is crucial to study the effectiveness of the system in helping students to achieve their future goals in a cost-efficient way. Since 96% of the cost is paid by society, it is also important to see if

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

students have the right incentives within the system to optimize total welfare (OECD, 2017). Belgium has a very high rate of grade retention in secondary education which comes at a large cost. The total cost of a year of study delay in Belgium amounts to at least $48,918/student or 11% of total expenditures on compulsory education, the highest rate in the OECD (OECD, 2013).

C Solution of the model

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

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

with $\gamma \approx 0.577$ the Euler constant and $\Phi_{it+1} = \Phi(x_{it+1})$ the choice set. $\bar{V}$ is used as an input in t (see (12)). First, students look for the optimal value of effort in every possible option in secondary education: y^*_{ijt}. As explained in section 2, an interior solution in the data is required and the following FOC should then be satisfied:

$$c_j(x_{it}, \nu_i) = \beta \sum_{\bar{g}} \frac{\partial \phi^{\bar{g}}_{ijt}(y_{it})}{\partial y_{it}} \bar{V}(x_{it+1}(\bar{g}), \nu_i) \text{ if } y_{it} = y^*_{ijt}. \tag{23}$$

As in the simple model of section 2, a sufficient condition to obtain an interior solution is to assume that students always believe there is a positive probability to avoid the worse performance outcome in any program. This avoids that $y^*_{ijt} = 0$. Furthermore, a positive marginal cost makes sure that it is never optimal to exert an infinite level of study effort. The FOC condition equalizes marginal costs and (expected) marginal benefits. As this does not depend on taste shocks ε or performance shocks η, it implies that students with the same state vector (x_{it}, ν_i) will choose the same effort levels in a given program: $y^*_{ijt} = y^*_j(x_{it}, \nu_i)$. In contrast to the simple model in section 2, I do not obtain a closed-form solution for $y^*_j(x_{it}, \nu_i)$. However, I can still estimate the optimal levels in the data (see section D for details about estimation). In counterfactual simulations, I run a grid search to find the new optimum (see section E for details about the simulations).

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

$$\Pr(d_{it} = j | x_{it}, \nu_i) = \frac{\exp(v_j(x_{it}, \nu_i, y_j^*(x_{it}, \nu_i)))}{\sum_{j' \in \Phi(x_{it})} \exp(v_{j'}(x_{it}, \nu_i, y_{j'}^*(x_{it}, \nu_i)))} \tag{24}$$

with v_{ijt} given by (12) for options in secondary education and (14) for options after secondary education. $\bar{V}(x_{it}, \nu_i)$ can also be calculated using:

$$\bar{V}(x_{it}, \nu_i) = \gamma + \ln \sum_{j \in \Phi(x_{it})} \exp(v_j(x_{it}, \nu_i, y_j^*(x_{it}, \nu_i))).$$

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

D Estimation details

I first explain the estimation of the model when the econometrician knows the type ν_i of every student and then allows this to be unobserved.

D.1 Higher education

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

$$
\begin{aligned}
\Psi_j^{HEE}(x_{it}, \nu_i) = \varphi_j^{HEE,0} & \\
& + S_i'(\varphi^{HEE,S,0} + \varphi^{HEE,S,level}\text{level\_HE}_j + \varphi^{HEE,S,STEM}\text{STEM}_j) \\
& + \nu_i'(\varphi^{HEE,\nu,0} + \varphi^{HEE,\nu,level}\text{level\_HE}_j + \varphi^{HEE,\nu,STEM}\text{STEM}_j) \\
& + \varphi^{HEE,\text{dist}}\text{distance\_HE}_{ij} \\
& + \tilde{d}'_{iT_i^{SE}}\varphi^{HEE,\text{SE}} \\
& + \text{delay}_{iT_i^{SE}}(\varphi^{HEE,delay,0} + \varphi^{HEE,delay,level}\text{level\_HE}_j + \varphi^{HEE,delay,STEM}\text{STEM}_j) \\
& + \varphi^{HEE,level\text{x}delay}\text{level\_SE}_{iT_i^{SE}} \times \text{delay}_{iT_i^{SE}} \\
& + X_{ij}'\varphi^{HEE,\text{interact}}
\end{aligned}
$$

Level_HE$_j$ is the level of the higher education program. I follow Arcidiacono (2005) and define the level for each type of higher education by the average math ability of the enrolling students. I use professional college as a benchmark (0.20) and calculate differences with academic college (0.59) and university (0.79). Distance_HE$_{ij}$ is the distance in kilometers from the student's home to the chosen option. $\tilde{d}_{iT_i^{SE}}$ is a vector of dummy variables for each possible program a student can graduate from in high school and delay$_{iT_i^{SE}}$ the years of accumulated study delay. Since there are few students in the academic track that do not enroll in higher education, I do not distinguish between elective courses and estimate a common effect of each track on enrollment in the benchmark professional college. I also include a vector of interactions X_{ij} that includes all interactions between characteristics of the high school program the student graduated in (academic level, intensive math, classical languages) and the characteristics of the higher education program (level and STEM major).

I impose a similar model for graduation from higher education. I use a similar functional form for $\Psi_j^{HED}(.)$ as I did for $\Psi_j^{HEE}(.)$, but I also add more interaction effects in X_{ij} to take into account the higher education enrollment decision.[34] In particular, I include dummy variables for choosing the same level, upgrading a level, and choosing the same major. I add a shock that is distributed extreme value type 1 such that I obtain logit probabilities. Since these shocks are iid, it is important to take into account the enrollment decision to capture the correlation between enrollment decisions and the final degree a student obtains.

D.2 Reduced forms of high school data

In section 2, I explained how a measure of performance can be used to back out the optimal level of effort. This is still possible in the current model and follows from the FOC (23). A first implication of this is that students with the same state vector will choose the same effort levels within each program. Let y^*_{ijt} be the optimal choice of y_{it}, conditional on program choice j. We can now substitute this in the definition of y_{it} (10):

$$y^*_{ijt} = \frac{1 - \Pr(g^{track}_{it+1} = 0 | d_{it} = j, x_{it}, \nu_i)}{\Pr(g^{track}_{it+1} = 0 | d_{it} = j, x_{it}, \nu_i)}$$

with the current grade deterministic in d_{it} and x_{it} and $y^*_{ijt} = y^*_j(x_{it}, \nu_i)$. Note that both x_{it} and ν_i are observed here, therefore y^*_{ijt} is easily obtained from the observed probability to obtain the lowest performance outcome in each j when students

[34] In contrast to college enrollment rates, there is sufficient variation in graduation rates within programs of the same track. Therefore, I do not need to restrict the common parameters of the effect of study programs to be the same.

behave optimally in the data. However, the finite number of observations and the large state space do not allow me to do this. Therefore, I recover the optimal levels and the performance thresholds by estimating an ordered logit model for the track performance outcome with index $\ln y_j^*(x_{it}, \nu_i)$ and cut points $\bar{\eta}_{jr}^{track}$. The functional form of the index is similar to what is imposed for the fixed cost parameters (see below, equation (26)), but I allow for more flexibility by letting each initial observed and unobserved characteristic be track-specific and change (linearly) over different grades. I also allow distance to higher education options to affect performance and I add an effect of the lagged study program (academic level and dummy variables for intensive math and classical languages). Note that some of the thresholds are not identified from the data but from the institutional context that imposes restrictions on mobility (i.e. some thresholds can be ∞). I allow the thresholds to differ not only by different programs but also by the grade a student is in. Because there is little variation in the data, I restrict the program-specific part through three parameters that capture differences in the increase in thresholds for obtaining a higher outcome in each track. This is then assumed to be constant over grades and tracks (see Table J14). The ordered logit model also generates the probabilities for each performance outcome. For elective courses, I use the predicted values of $\ln y_j^*(x_{it}, \nu_i)$ and estimate the specification in equation (11). Both can then be used to construct the joint probabilities $\phi_{ijt}^{\bar{g}}(y_j^*(x_{it}, \nu_i))$.

As in Arcidiacono et al. (2023), I also obtain predicted values of $\Pr(d_{it}|x_{it}, \nu_i)$ (the CCPs) by estimating a flexible conditional logit with an index, similar to the index I used to predict effort. I assume a functional form that is linear in observed and unobserved characteristics for each student characteristic, and I allow for more flexibility than in fixed costs by letting them be track-specific and change linearly over different grades. I also allow distance to higher education options to affect choices, while they are excluded from fixed costs. As explained further, the CCPs will be used to avoid solving the model during estimation and to back out the unobserved types in a first stage.

D.3 Cost estimates

The FOC (23) allows us to write the conditional value functions without an unknown marginal cost function. Substituting the utility function (13) in the conditional value function (12), after substituting marginal costs by (23) gives:

$$
\begin{aligned}
& v_j(x_{it}, \nu_i, y_j^*(x_{it}, \nu_i)) \\
& = -C_j^0(x_{it}, \nu_i) \\
& + \beta \sum_{\bar{g} \in G} \left[\bar{V}(x_{it+1}(\bar{g}), \nu_i) \left(\phi_{ijt}^{\bar{g}}(y_j^*(x_{it}, \nu_i)) - \left. \frac{\partial \phi_{ijt}^{\bar{g}}(y_{it})}{\partial y_{it}} \right|_{y_{it} = y_j^*(x_{it}, \nu_i)} y_j^*(x_{it}, \nu_i) \right) \right].
\end{aligned} \tag{25}
$$

We already recovered $y_j^*(x_{it}, \nu_i)$ and $\phi_{ijt}^{\bar{g}}(y_j^*(x_{it}, \nu_i))$ from the data. $\frac{\partial \phi_{ijt}^{\bar{g}}(y_{it})}{\partial y_{it}}$ can be derived from the distributional assumptions on the performance measure. As explained in the text, it is the product of three ordered logit probabilities. We can apply the chain rule, knowing that for each ordered logit model we can find the derivative with respect to y_{it} recursively:

$$
\begin{aligned}
\frac{\partial \Pr(g_{it}^a = 0 | d_{it}, x_{it}, \nu_i, y_{it})}{\partial y_{it}} &= -\alpha_y^a \frac{1}{y_{it}} \Pr(g_{it}^a = 0 | d_{it}, x_{it}, \nu_i, y_{it})(1 - \Pr(g_{it}^a = 0 | d_{it}, x_{it}, \nu_i, y_{it})) \\
\frac{\partial \Pr(g_{it}^a = \bar{g} | d_{it}, x_{it}, \nu_i, y_{it})}{\partial y_{it}} &= -\alpha_y^a \frac{1}{y_{it}} \left(\Pr(g_{it}^a \leq \bar{g} | d_{it}, x_{it}, \nu_i, y_{it}) \Pr(g_{it}^a > \bar{g} | d_{it}, x_{it}, \nu_i, y_{it})\right) \\
&\quad - \sum_{\tilde{g} < \bar{g}} \frac{\partial \Pr(g_{it}^a = \tilde{g} | d_{it}, x_{it}, \nu_i, y_{it})}{\partial y_{it}} \text{ for } \bar{g} > 0
\end{aligned}
$$

with $(a = track, clas, math)$ and $\alpha_y^{track} = 1$.

After solving the model for $\bar{V}(x_{it+1}(\bar{g}), \nu_i)$, we can use the logit probabilities (24) with these conditional value functions to estimate the value of a degree μ^{degree} and a specification for fixed costs $C_j^0(.)$ by using maximum likelihood. I assume the following functional form:[35]

$$
\begin{aligned}
C_j^0(x_{it}, \nu_i) = {} & \mu_j^0 + \mu_j^{\text{grade}} \text{grade}_{ijt} \\
& + S_i'(\mu_j^{S,0} + \mu^{S,\text{level}} \text{level\_SE}_{ijt} + \mu^{S,\text{math}} \text{math}_{ijt} + \mu^{S,\text{clas}} \text{clas}_{ijt}) \\
& + \nu_i'(\mu_j^{\nu,0} + \mu^{\nu,\text{level}} \text{level\_SE}_{ijt} + \mu^{\nu,\text{math}} \text{math}_{ijt} + \mu^{\nu,\text{clas}} \text{clas}_{ijt}) \\
& + \mu_{time} \text{time}_{ijt} \\
& + \text{retention}_{ijt}'(\mu^{\text{ret},0} + \mu^{\text{ret},\text{level}} \text{level\_SE}_{ijt}) \\
& + \mu_{up} \text{upgrade}_{ijt} + \mu_{down} \text{downgrade}_{ijt} \\
& + \mu_{staymath} \text{math}_{ijt} \times \text{math}_{it-1} + \mu_{stayclas} \text{clas}_{ijt} \times \text{clas}_{it-1}.
\end{aligned} \tag{26}
$$

μ is a vector of parameters to estimate. S_i is a vector of time-invariant observed student characteristics, ν_i is a vector of dummy variables that indicate to which type the student belongs, time_{ijt} is the daily commuting time to the closest school that offers the study program in the current grade and grade_{ijt} is the grade a student is in (set such that 1 is the first year of high school). Level_SE_{ijt} is the academic level of the track a student is in with 0 the vocational track, 1 the middle-practical track, 2 the middle-theoretical track, and 3 the academic track and math and clas refer to respectively programs with intensive math and with classical languages. Grade retention is captured by the 2x1 vector: retention_{ijt}. This vector contains a flow variable: a dummy equal to one if the student is currently in the same grade as the year before ("Repeat") and a stock variable that captures the years of study delay accumulated in previous years ("Study delay"). Finally, upgrade_{ijt} and downgrade_{ijt} are dummy variables indicating if a student is currently in a track with at a higher or lower academic level than the year before and $\mu_{staymath}$ and $\mu_{stayclas}$ capture preferences to stay in a program with the same elective courses.

Note that in section 4, the scale of the utility function was implicitly normalized

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

to unity. Therefore, all parameters μ are identified. However, to directly interpret the cost estimates, I rescale the parameters by dividing them by μ_{time}. This way, the cost estimates can be measured in daily commuting time.

Finally, marginal costs $c_j(.)$ can be recovered from the FOC (23) without imposing additional structure.

D.4 CCP estimation

Hotz and Miller (1993) introduced the CCP method as an alternative to solving dynamic models, which is particularly useful if there is a terminal action (Arcidiacono and Ellickson, 2011). Hotz and Miller (1993) show that the future value term can be written as the conditional value function of an arbitrary choice and a nonnegative correction term that depends on its probability in the data:

$$\bar{V}(x_{it+1}, \nu_i) = \gamma + v_{d^*}(x_{it+1}, \nu_i, y_j^*(x_{it}, \nu_i)) - \ln \Pr(d_{it+1}^* | x_{it+1}, \nu_i) \tag{27}$$

with $\gamma \approx 0.577$ the Euler constant, d_{it+1}^* an arbitrary option $j = d^*$ and $v_{d^*}(.)$ the conditional value function of this option.

Case 1: $j = 0$ available in $t + 1$

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 (14) in (27), with $\Psi_0^{HEE}(.) = 0$:

$$\overline{V}(x_{it+1}, \nu_i) = \gamma + \text{Degree}_{it}' \mu^{\text{degree}} - \ln \Pr(d_{it+1}^0 = 1 | x_{it+1}, \nu_i). \tag{28}$$

We can now substitute (28) in (25), such that for all $j \in se$:

$$
\begin{aligned}
& v_j(x_{it}, \nu_i, y_j^*(x_{it}, \nu_i)) && (29) \\
&= -C_j^0(x_{it}, \nu_i) + \beta\gamma \\
&+ \beta \sum_{\bar{g} \in G} \left[\begin{array}{c} \left(\text{Degree}_{it}'(\bar{g})\mu^{\text{degree}} - \ln \Pr(d_{it+1} = 0 | x_{it+1}(\bar{g}), \nu_i)\right) \\ \left(\phi_{ijt}^{\bar{g}}(y_j^*(x_{it}, \nu_i)) - \left.\frac{\partial \phi_{ijt}^{\bar{g}}(y_{it})}{\partial y_{it}}\right|_{y_{it} = y_j^*(x_{it}, \nu_i)} y_j^*(x_{it}, \nu_i) \right) \end{array} \right].
\end{aligned}
$$

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 4.5). As in Hotz and Miller (1993), a nonparametric estimate of $\Pr(d_{it+1} = 0 | x_{it+1}, \nu_i)$ can be recovered from the data before estimating the model.

These conditional value functions can now be used as inputs in logit probabilities to recover the fixed cost parameters without having to solve the model.

Case 2: $j = 0$ available in $t + \rho_{it}$

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: $\rho_{it} = \max\{1, 18 - Age_{it}\}$. We now need to repeat the CCP method in future values until the outside option is available. This is an application of finite dependence, introduced in Arcidiacono and Miller (2011). In contrast to their application on problems that have a renewal action in the future, I apply it to the terminal action of choosing to leave secondary education in the outside option (no higher education). The exposition in this section is similar to Arcidiacono and Miller (2011) and Arcidiacono and Ellickson (2011).

The choice probabilities (24) at the optimal levels of the effort can be written by using differenced value functions. Let $v_j^*(x_{it},\nu_i) \equiv v_j(x_{it},\nu_i,y_j^*(x_{it},\nu_i))$ be the conditional value function at the optimal level of effort and $u_j^*(x_{it},\nu_i) \equiv u_j(x_{it},\nu_i,y_j^*(x_{it},\nu_i))$ the flow utility at this level:

$$\Pr(d_{it}=j|x_{it},\nu_i) = \frac{\exp\left(v_j^*(x_{it},\nu_i) - v_{j'}^*(x_{it},\nu_i)\right)}{1+\sum_{j^\circ\in\Phi(x_{it})}\exp\left(v_{j^\circ}^*(x_{it},\nu_i) - v_{j'}^*(x_{it},\nu_i)\right)}$$

$$\begin{aligned}
&\text{with } v_j^*(x_{it},\nu_i) - v_{j'}^*(x_{it},\nu_i) \\
&= u_j^*(x_{it},\nu_i) + \beta\sum_{\bar{g}\in G}\phi_{ijt}^{\bar{g}}(y_{ijt}^*)\bar{V}(x_{it+1}(\bar{g})) \\
&- u_{j'}^*(x_{it},\nu_i) - \beta\sum_{\bar{g}\in G}\phi_{ij't}^{\bar{g}}(y_{ij't}^*)\bar{V}(x_{it+1}(\bar{g})),
\end{aligned} \tag{30}$$

for any $j^{'} \in \Phi(x_{it})$. Substitute the CCP representation of the future value as a function of the CCP of an arbitrary choice and its conditional value function (27) in (30):

$$\begin{aligned}
&v_j^*(x_{it},\nu_i) - v_{j'}^*(x_{it},\nu_i) \\
&= u_j^*(x_{it},\nu_i) + \beta\sum_{\bar{g}\in G}\phi_{ijt}^{\bar{g}}(y_{ijt}^*)\left(\gamma + v_{d^*}^*(x_{it+1}(\bar{g}),\nu_i) - \ln\Pr(d_{it+1}^*|x_{it+1}(\bar{g}),\nu_i)\right) \\
&- u_{j'}^*(x_{it},\nu_i) - \beta\sum_{\bar{g}\in G}\phi_{ij't}^{\bar{g}}(y_{ij't}^*)\left(\gamma + v_{d^*}^*(x_{it+1}(\bar{g}),\nu_i) - \ln\Pr(d_{it+1}^*|x_{it+1}(\bar{g}),\nu_i)\right).
\end{aligned} \tag{31}$$

Define the cumulative probability of being in a particular state given the current state variable and choice, and a particular decision sequence $d_i^* = (d_{it}, d_{it+1}^*, d_{it+2}^*, ...d_{it+\rho_{it}}^*)$:

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

with $\phi^{\bar{g}}_{id^*\tau}(y^*_{d^*}(x_{i\tau}, \nu_i))$ the probability of receiving performance outcome $\bar{g}$ at time $t = \tau + 1$, in the program a student will be at $t = \tau$ 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: $d^{'}_i = (d'_{it}, d^*_{it+1}, d^*_{it+2}, ...d^*_{it+\rho_{it}})$.[36] We can then repeat the CCP method in each of the future periods and rewrite (31) as the sum of future flow utilities and CCPs until the outside option becomes available at $t + \rho_{it}$:

$$
\begin{aligned}
& v^*_j(x_{it}, \nu_i) - v^*_{j'}(x_{it}, \nu_i) \\
& = u^*_j(x_{it}, \nu_i) - u^*_{j'}(x_{it}, \nu_i) \\
& + \sum_{\tau=t+1}^{t+\rho_{it}-1} \beta^{\tau-t} \sum_{\bar{g}\in G} [u^*_{d^*}(x_{i\tau}(\bar{g}), \nu_i) - \ln \Pr(d^*_{i\tau}|x_{i\tau}(\bar{g}), \nu_i)]\kappa^*_{\tau-1}(\bar{g}|x_{it}, \nu_i) \\
& - \sum_{\tau=t+1}^{t+\rho_{it}-1} \beta^{\tau-t} \sum_{\bar{g}\in G} [u^*_{d^*}(x_{i\tau}(\bar{g}), \nu_i) - \ln \Pr(d^*_{i\tau}|x_{i\tau}(\bar{g}), \nu_i)]\kappa'_{\tau-1}(\bar{g}|x_{it}, \nu_i) \\
& + \beta^{\rho_{it}} \sum_{\bar{g}\in G} \overline{V}(x_{t+\rho_{it}}(\bar{g}), \nu_i)\kappa^*_{t+\rho_{it}-1}(\bar{g}|x_{it}, \nu_i) \\
& - \beta^{\rho_{it}} \sum_{\bar{g}\in G} \overline{V}(x_{t+\rho_{it}}(\bar{g}), \nu_i)\kappa^{'}_{t+\rho_{it}-1}(\bar{g}|x_{it}, \nu_i).
\end{aligned}
$$

$\overline{V}(x_{t+\rho_{it}}, \nu_i)$, the value of behaving optimally when the outside option is available and can be written as in (28). The calculation of the value function is now possible after choosing the arbitrary options in each period, the prediction of their CCPs, and the predictions of optimal effort in the study program. However, further simplifications follow from a good choice of "arbitrary" options.

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

[36] 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.

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 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 $u_j^*(x_{it}, \nu_i) = -C_j^0(x_{it}, \nu_i) - c_j(x_{it}, \nu_i) y_j^*(x_{it}, \nu_i)$. By replacing the marginal cost of effort with the marginal benefit of effort in the data, future value terms also enter directly into $u_j^*(x_{it}, \nu_i)$ (see (25)). Because $\sum_{\bar{g} \in G} \frac{\partial \phi_{ijt}^{\bar{g}}(y_{it})}{\partial y_{it}} = 0$, all terms that do not depend on performance drop out such that the same simplifications arise because of finite dependence.

D.5 Unobserved heterogeneity

To allow for types to remain unobserved to the econometrician, I follow the two-stage estimator of Arcidiacono and Miller (2011). 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_i^0. 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_i^0, we can allow for this correlation.[37] The loglikelihood function is

$$\ln L_i = \ln \sum_{m=1}^{M} \pi_{m|age_i^0} L_i^m$$

with

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

with $L_{it}^{program,m}$ and $L_i^{HEE,m}$ given by logit choice probabilities (24), with conditional value functions (29) and (14). $L_i^{HED,m}$ is given by the conditional logit probabilities on the different possibilities for higher education graduation outcomes. The likelihood contribution of the performance outcome in secondary education is given by ordered logit probabilities $L_{it+1}^{\text{performance},m}$ and $L_{it}^{ccp,m}$ are the CCP predictors. Note that the inclusion of unobserved types makes the function no longer additively separable such that sequential estimation is not possible.

Arcidiacono and Miller (2011) show that additive separability can be restored. The estimation procedure is an adaptation of the EM algorithm. It starts from a random probability of each observation to belong to each type. The entire model can then be estimated as explained above but weighs each observation-type combination by the probability that the student belongs to the type. Afterward, the joint likelihood of the data conditional on each type is used to update the individual type probabilities, conditional on the data, using Bayes rule. This is repeated until convergence of the likelihood function. I use the two-stage estimator of Arcidiacono and Miller (2011) which implies that in the calculation of the joint likelihood, reduced form estimates of the CCPs are used for $L_{it}^{program,m}$, instead of the choice probabilities from the structural model. This means that the fixed cost parameters and the common

[37] 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.

component of the value of a degree are recovered in the second stage. Finally, the FOC (23) is used to recover the marginal costs.

Standard errors are obtained using a bootstrap procedure. I sample students with replacement from the observed distribution of the data and use 150 replications. Since the EM algorithm takes some time to converge, I do not correct for estimation error in the probabilities to belong to each type.

D.6 Calculation of ATT

The ATTs are calculated as follows:

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

with $E_{x,\nu}$ an expectations operator over the empirical distribution of the observables x and the estimated distribution of the unobserved types ν. P_j^{HE} is the probability of the higher education outcome (enrollment or graduation) as a function of the state variables. $x_{it_{HE}}(j')$ is the observed state vector of student i in the data at the time the outcome is realized $t = t_{HE}$ and $x_{it_{HE}}(j^0)$ is the same vector but with the graduation track replaced by an arbitrary benchmark program j^0. The ATT then calculates the average effect on HE of graduating high school in j' instead of j^0 for the group of students who graduated from j' in the data.

E Simulation details

All predicted values are calculated as follows. I first categorize students by their demographic characteristics: gender, language ability, math ability, SES, and the age they start high school. I discretize the observed ability distribution by creating four equally sized groups for each measure. Every student then belongs to one group which is a unique combination of these variables. Within each group, I use the average travel times and distances. Each group is then used to calculate the value functions for each unobserved type. To limit the number of calculations, I drop groups with less than 10 students and verify that this has a negligible effect on the distribution of student characteristics.

E.1 High school

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

This procedure allows for a substantial total number of draws while needing only a limited number of students to use for a grid search to find the optimal effort level within each possible program. The grid search for effort levels starts at the optimal value of the scalar y_{it} in each program j in the data: $y_j^*(x_{it}, \nu_i)$ and looks for better levels using five sequential loops and an additional step to check for a corner solution. The first loop looks at changes in the log of effort by 1 unit with a minimum of -5 and a maximum of +5. The second loop divides steps and thresholds by five, the third by 25, the fourth by 125, and the fifth by 625, such that the final precision is 0.0016 (which is about 0.16% for effort y). Finally, I check if a corner solution is optimal by setting $y = 0$ and changing the performance distribution to predict the worse outcome with probability 1.

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

E.2 Higher education

To evaluate the impact of counterfactuals that affect students during secondary education, a structural model for the decision in high school is needed as the same cost parameters and values of degree, will have new implications for optimal program and effort choices. To predict the policy impact after secondary education, we only need to know how they are influenced by high school outcomes, after controlling for observed and unobserved student characteristics. Therefore, I model a reduced form function only. This is similar to the approach in the dynamic treatment effect literature (Heckman et al., 2016), but I only apply it to choices after leaving high school to be able to do counterfactual simulations during secondary education in which students are forward-looking.

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

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

with $E_{x,\nu}$ an expectations operator over the empirical distribution of the observables x and the estimated distribution of the unobserved types ν. P_j^{HE} is the probability of the enrollment decision or higher education degree outcome of each college option as a function of the state variables.

E.3 Fit of the model

Table E1 shows the ability of the model to replicate the actual data. The model does a good job of predicting the patterns in the data such that it can be used for counterfactual simulations. We see that graduation rates in different track and higher education outcomes are predicted very precisely. There is a slight overprediction in the number of students with a B-certificate leading to a small overprediction in the number of students with study delay.

Table E1: Predictions of the model

		Data	Predictions	
High school (% of students)				
Academic		38.27	40.02	(2.07)
	clas+math	5.06	5.03	(0.65)
	clas	6.11	3.18	(0.42)
	math	13.24	14.59	(1.27)
	other	13.86	17.22	(1.29)
Middle-Theoretical		15.86	16.10	(1.24)
	math	2.42	3.11	(0.46)
	other	13.44	12.99	(0.97)
Middle-Practical		11.85	8.14	(1.19)
Vocational		19.43	21.57	(0.89)
Dropout		14.60	14.17	(0.67)
Students with at least 1 B-certificate		35.40	37.53	(0.81)
Students with at least 1 C-certificate		30.01	30.69	(0.77)
Students with at least 1 year of study delay		31.62	33.22	(0.91)
Higher education (% of students)				
Enrollment		58.18	58.15	(0.75)
Graduation		44.01	44.25	(0.75)
SES gap at graduation		38.49	39.73	(1.13)
University degree		12.43	11.22	(0.55)
Academic college degree		6.05	6.26	(0.38)
Professional college degree		25.53	26.77	(0.69)
Degree in STEM major		17.76	18.01	(0.65)

Note: Clas= classical languages included. Math= intensive math. Observed outcomes in the data and predictions from the dynamic effort model. SES gap at graduation: difference in percentage college graduates between high and low SES. Bootstrap standard errors of predicted values in parentheses.

E.4 Welfare

Opportunity cost

I assume an opportunity cost of \$10/hour. This is chosen to approximate the opportunity cost of students in high school and is consistent with Kapor et al. (2020). Students are not allowed to work until they are 15 years old and the wage often depends on their age. In 2012 the minimum wage ranged between ¤6.8 and ¤9.7/hour.[38] Only a small amount of taxes is paid on this if they work a limited amount of hours. To compare to OECD estimates, I use the PPP-adjusted exchange rate of dollars (0.82), which results in wages between \$8 and \$12. Note that the model is in years while the estimates are scaled in minutes/day. Therefore, I multiply them by the wage per minute (\$10/60) and the 177 school days there are in a year.

Gains from reducing grade retention

The direct cost and the total foregone earnings can be found in Table IV.1.6 in OECD (2013). The direct cost of a student who repeats a grade is \$9,713. The downgrade policy decreases grade retention rates by 9.82 % points and therefore generates a government saving of \$950 per student. To only capture the externality in foregone earnings, I subtract the net income (49%). This number was calculated by dividing column (7) by column (1) in Table A10.2 in OECD (2012). The externality in the downgrade policy then amounts to \$1,960 per student.

Reinvestment of gains

Estimates in the literature for the effect of a one-time "helicopter drop" increase of \$1,000 on the ability distribution are around 1% to 2% of a standard deviation (Gigliotti and Sorensen, 2018; Lafortune et al., 2018). This implies that reinvesting the efficiency gains of the "Downgrade" policy could result in substantial gains for students. Using the estimates in Table J20 and the savings from avoiding grade retention (\$2,910), a 1.5% effect per \$1,000 on each of the observed ability measures in

[38] https://www.jobat.be/nl/artikels/wat-is-het-minimumloon-voor-een-jobstudent/ (consulted March 2018).

the downgrade policy would bring back $1,200 in student welfare, increase graduation rates in higher education by 1.3 % points, reduce study delay by 0.4 % points and dropout by 0.6 % points. This in turn also creates additional savings that could be reinvested.

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

E.5 Alternative policy changes

As student welfare decreases, I also investigate alternative policies that are guaranteed to increase student welfare. First, I consider the impact of lowering the compulsory schooling age from 18 to 15 years old. The reason is that underperforming students might prefer to leave school rather than accumulate study delay. While high school dropout is costly for society, if these students would leave anyway, it could result in a situation where everyone is better off. Second, I allow more flexible switching between tracks. Students can currently only switch downward, except for the first two grades in which they can also switch upwards between academic and middle tracks (although at a high estimated cost). The reasoning behind this is that they would be suboptimally prepared if they move upwards. I simulate the impact of allowing them to upgrade.[39]

The results can be found in Table E2. While the repeat and downgrade policies increase the control of teachers (or the school system) on the students' path, these

[39] I make it possible to obtain the certificates that make them qualify for higher tracks, remove the impact on marginal costs of switching track and reducing the fixed cost of upward mobility to the level of the fixed cost of downward mobility (as some persistence is likely to still occur, e.g. to stay in the same class as their friends).

policies rather give more flexibility to students (or their parents). By construction, this generates more student welfare. Lowering the age of compulsory schooling increases average welfare by \$ 880, allowing upward mobility increases it by \$ $1,710$. However, the risk of more flexibility is that it amplifies negative externalities coming from an undervaluation of the cost of grade retention, dropout and incompletion of higher education. While lowering the age of compulsory schooling decreases study delay, it not only affects those that would drop out anyway. Dropout increases by a substantial 12 % points. Allowing for upward mobility decreases graduation from the vocational track, but the switch is not to other tracks, instead we see an increase in dropout of 5.0 % points, as well as an increase in study delay of 8.4 % points. The surprising result can be explained when we look at the decomposition of the welfare effects, as well as the impact on certificates. Most of the welfare is generated from students gaining from idiosyncratic taste shocks on period utility. I.e. several students in a lower track are drawn by a temporary reason to attend a higher track. Once there, they perform poorly. The main issue is that this is not only reflected in B-certificates (which would still allow them to transition to lower tracks later on), but we also see an 11 %point increase in C-certificates, which requires repeating grades or dropping out. The reason they perform more poorly can be seen in the marginal cost estimates (Table 2). As explained in section 5, low-ability students face higher marginal costs, especially when they attend higher-level tracks. This means that attending a high track to obtain a certificate that allows a lower track is relatively more difficult for lower ability students than performing well in that low track. This reflects a mismatch between the ability of the student and the academic level of the program.

It is important to note that this is not suggesting upward mobility should not be applied. It is successful in Germany (Dustmann et al., 2017), partly through the creation of specialized high schools for upgrading students. The simulations rather show that upward mobility should be accompanied by other measures to be a success.

For example, it can be restricted to students with reasonable chances of success, or their success can be directly influenced to reduce the increase in marginal costs many upgraders experience.

Table E2: Alternative policy changes

	Status quo	Policy change B-certificate			
		Comp. schooling until 15		Facilitate upgrading	
Panel A: educational outcomes	*%*	*Change in % points*			
High school					
Academic track	40.02	-0.25	(0.07)	0.67	(0.34)
Middle-theoretical track	16.10	-0.90	(0.13)	-0.13	(0.25)
Middle-practical track	8.14	-1.36	(0.22)	0.71	(0.28)
Vocational track	21.57	-9.87	(0.61)	-6.27	(0.51)
Dropout	14.17	12.38	(0.82)	5.02	(0.46)
At least 1 B-certificate	37.53	0.25	(0.15)	11.41	(0.64)
At least 1 C-certificate	30.69	-2.15	(0.33)	10.77	(0.96)
At least 1 year of study delay	33.22	-4.91	(0.43)	8.37	(0.87)
Higher education					
Enrollment	58.15	-2.12	(0.23)	-0.17	(0.19)
Graduation	44.25	-0.79	(0.10)	-0.69	(0.15)
SES gap at graduation	39.73	0.49	(0.80)	-0.19	(0.18)
Panel B: student welfare		*Change in $1000*			
Total student welfare		0.88	(0.14)	1.71	(0.25)
Fixed costs (-)		-1.73	(0.29)	1.81	(0.26)
Variable costs (-)		-0.71	(0.09)	-0.15	(0.04)
Expected payoff after high school (+)		-0.56	(0.13)	-0.68	(0.11)
Taste shocks (+)		-0.99	(0.16)	4.05	(0.54)

Note: Predictions from the dynamic effort model. C-certificate: repeat grade. B-certificate = students acquired skills to proceed to next grade but only if they downgrade, i.e. switch to track of lower academic level or drop an elective course. Status quo = students can choose to downgrade or repeat grade after obtaining a B-certificate, Comp. schooling until 15: full time drop out option available in the year in which the student turns 15 instead of 18. Facilitate upgrading: the student can obtain certificates that make them qualify for higher tracks, remove the impact on marginal costs of switching track and reduce fixed cost of upward mobility to the level of the fixed cost of downward mobility. SES gap at graduation: difference in percentage college graduates between high and low SES. Opportunity cost of time: $10/h. Bootstrap standard errors in parentheses.

F Sensitivity analysis

This Appendix discusses the sensitivity checks. First, we show evidence that the high school outcomes we include in the model capture well the human capital accumulated in high school. Second, we look at the impact of changes to the model on the main results from the counterfactual simulations.

F.1 Relevance high school outcomes

The model implies that changes in effort affect the future through their impact on observable states: it affects study delay and the choice set over programs in the next grade. Because higher education outcomes depend on the graduation track and study delay, it gives students an incentive to exert effort. This implies that our measure of effort cannot capture an increase in human capital, above what is captured by the graduation track and the years of study delay. To check the sensitivity to this assumption, we make use of test score data obtained at the end of grade 10 and the end of grade 12, but excluded from the model. There are several reasons to exclude this from the model. The first is to keep the state space small, which facilitates estimating and solving the model. This is especially a concern because of the continuity of the test score data. A second reason to exclude it is due to attrition. This is caused by drop out, more than one year of study delay, or students moving to schools where the tests did not take place. Finally, the expected impact of the test score in this context is expected to be limited. While any test score could capture human capital improvements, they do not have any signaling value here. First, there are no admission restrictions in college. Therefore, a higher test score does not help to enter a program. Second, Flanders does not do standardized testing. The measures we have, are obtained for the sole purpose of research on this sample of students.

To investigate the impact of human capital accumulation beyond what we include in the model, we re-estimate higher education enrollment and graduation with the additional test score variables. Note that we are not interested in their effect, or the effect of high school background, conditional on these measures. An important part of the effect of tracks and study delay is coming from its impact on human capital, likely captured by these additional test scores. This creates a "bad control" problem when directly interpreting them (Angrist and Pischke, 2009). What we are interested in, is the change in predicted probabilities by including these measures. If there are large changes, it tells us that we are ignoring an important dimension of human capital accumulation in the model, casting doubts on the model being able to capture changes in effort.

We start by including only the tests for math and language skills taken at the end of grade 10. This reduces the sample from 5158 to 4180 students. We do the same for the end of grade 12, but this creates a much more selected sample of 2600 students. We keep the individual-level type probabilities fixed and re-estimate high school enrollment and graduation. We then summarize the difference in predicted probabilities with and without including the test scores. The results can be found in Table F1. We see that adding the test score data hardly increases the standard deviation of predictions. Also at the individual levels, we see very little changes. The most extreme 10% of changes are still only between 1 and 3 % points. This suggests that there is little to gain from including test scores in the model, meaning that effort beyond what we include in the model is likely of little importance.

To show that this is not simply explained by the importance of initial conditions (i.e. characteristics of students before they enter high school), I also show the impact of removing the high school background (accumulated study delay and the type of high school degree). Indeed, this has a much bigger impact on the predicted probabilities, especially if we look at the tails of the distribution where we see differences of up to 15 % points.

Table F1: Sensitivity analysis: adding test scores

	Predictions		Difference with baseline				
	Mean	SD	Mean	SD	p10	Median	p90
Sample with test scores of grade 10 (N=4180)							
Enrollment higher education							
Baseline	0.64	0.40					
Add recent test scores	0.64	0.40	0.00	0.03	-0.02	0.00	0.02
Remove impact high school	0.64	0.37	0.00	0.14	-0.15	0.00	0.15
Graduation higher education							
Baseline	0.50	0.41					
Add recent test scores	0.50	0.41	0.00	0.02	-0.01	0.00	0.01
Remove impact high school	0.50	0.40	0.00	0.09	-0.09	0.00	0.08
Sample with test scores of grade 12 (N=2600)							
Enrollment higher education							
Baseline	0.89	0.17					
Add recent test scores	0.89	0.17	0.00	0.02	-0.02	0.00	0.02
Remove impact high school	0.89	0.15	0.00	0.08	-0.12	0.00	0.08
Graduation higher education							
Baseline	0.71	0.31					
Add recent test scores	0.71	0.31	0.00	0.04	-0.03	0.00	0.03
Remove impact high school	0.71	0.29	0.00	0.09	-0.11	0.00	0.09

Note: Predictions, conditional on the observed situation in the period before and using the estimated, individual-level type-probabilities. "Baseline" uses the same equations as in the paper. "Remove impact high school" removes the effects of study delay and type of high school degree (i.e. final study program). "Recent test scores" adds the test score data of grade 10 or 12 in the index that predicts higher education outcomes: an impact on higher education, and an interaction with the level and major. The sample size in the paper is 5158 but is restricted here because of attrition when using more recent test score data.

F.2 Model specifications

This subsection discusses the impact of alternative model specifications on the counterfactual simulations of this paper.

Table F2 looks at the impact of allowing for observed and unobserved ability. Compared to the baseline, we see important differences after allowing for heterogeneity by ability, especially on predicted higher education graduation. In the "Repeat" policy we see an underestimation of the decrease in graduating from higher education (-0.61 instead of -1.70 % points), while in the "Downgrade" policy we see an overestimation of the decrease (-0.84 instead of -0.30 % points). These results can be explained by a failure to take into account the ability bias on the estimated effect of tracks. When isolated, both observable ability measures and unobserved types move the estimate closer to the baseline results. However, for the "Repeat" policy it is mainly coming from the inclusion of observable measures of ability, while for the downgrade policy the types help more. Note however that the ability bias was also smaller in the downgrade policy. These results can be explained by the nature of the data. The availability of rich, continuous measures of ability helps a lot to capture the main source of ability bias. This leaves room for unobserved types to capture more subtle differences between students. Therefore, the limiting structure of having a finite number of types becomes a smaller concern with rich data. This is also confirmed by adding a third unobserved type to the model, which changes little to the main results.

Table F3 compares the baseline estimation method with two approaches that use different identifying assumptions. As discussed in section 4.7, travel time to high school programs is excluded from equations that predict higher education enrollment and graduation but exclusion is not required for identification (Heckman and Navarro, 2007). I therefore also estimate a specification in which I add measures of travel time to several programs in the final grade of high school. I add time to the vocational track, and differences in travel time for moving up a track for every other track. I also add the difference in travel time between options with and without intensive math and with and without classical languages. These travel times are interacted in the same way as other observable student characteristics. The resulting simulations are almost identical to the baseline results.

In a final specification, I follow Carneiro et al. (2003), Heckman et al. (2016), and Lin (2020) and use additional measurement data to identify unobserved heterogeneity. To do this, I add (ordered) logit models to the likelihood function of stage 1 of the estimation approach to predict the variables listed in Table F4. I use a discretized measure of students' IQ when they enter high school, as well as answers by their last

Table F2: Sensitivity analysis: observed and unobserved ability

	Study delay		**High school dropout**		**Higher education graduation**		**Student welfare**	
				Predicted value in %				
Status quo								
Baseline	33.22	(0.91)	14.17	(0.67)	44.25	(0.75)		
Obs ability: NO, types: NO	31.29	(1.41)	11.43	(1.37)	50.87	(3.75)		
Obs ability: YES, types: NO	33.93	(0.98)	15.24	(0.76)	43.78	(1.37)		
Obs ability: NO, types: YES	31.11	(1.11)	14.62	(0.64)	46.70	(1.76)		
Baseline + third type	33.33	(1.48)	13.48	(0.58)	44.62	(0.96)		
				Change in % points			*Change in $1000*	
Repeat policy								
Baseline	9.48	(0.57)	3.94	(0.33)	-1.70	(0.22)	-2.14	(0.26)
Obs ability: NO, types: NO	10.64	(0.70)	4.67	(0.55)	-0.61	(0.57)	-2.30	(0.27)
Obs ability: YES, types: NO	10.42	(0.59)	4.44	(0.37)	-1.80	(0.33)	-2.14	(0.24)
Obs ability: NO, types: YES	9.89	(0.61)	4.20	(0.38)	-0.72	(0.41)	-2.22	(0.26)
Baseline + third type	9.21	(0.50)	3.43	(0.30)	-1.65	(0.21)	-2.00	(0.19)
Downgrade policy								
Baseline	-9.82	(0.55)	-1.61	(0.25)	-0.30	(0.18)	-1.02	(0.14)
Obs ability: NO, types: NO	-9.16	(0.70)	-1.13	(0.38)	-0.84	(0.32)	-0.69	(0.11)
Obs ability: YES, types: NO	-10.35	(0.78)	-1.46	(0.34)	-0.57	(0.24)	-0.96	(0.14)
Obs ability: NO, types: YES	-9.22	(0.65)	-1.83	(0.24)	-0.26	(0.19)	-0.72	(0.11)
Baseline + third type	-10.31	(0.64)	-1.67	(0.23)	-0.16	(0.18)	-0.98	(0.12)

Note: Predictions from the dynamic model under alternative specifications. Obs ability refers to the variables on initial math and language ability. If types = YES, it means that two unobserved types are allowed for in the estimation. Baseline has both observed ability and types. Status quo = students can choose to downgrade or repeat grade after obtaining a B-certificate, Repeat = students must repeat grade after obtaining a B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining a B-certificate. Bootstrap standard errors in parentheses.

teacher in elementary school to questions that indicate levels of conscientiousness, extraversion, and agreeableness.[40] Furthermore, I add parental reports of their income category and work situation around the same time. Table F3 shows that adding these measures does not have an important impact on the counterfactual simulations. Furthermore, they give more insights into the nature of unobserved heterogeneity (see Table F5 and Table F6).[41] I find that unobserved types are important in capturing non-cognitive skills. The impact on IQ, parental income, and work situation has

[40] I use the same questions as in Shure (2021).

[41] Note that the model was re-estimated and therefore the types can change identity. Indeed, type 2 is now the high ability type while that was type 1 in the baseline specification.

Table F3: Sensitivity analysis: identification unobserved ability

	Study delay		**High school dropout**		**Higher education graduation**		**Student welfare**	
			Predicted value in %					
Status quo								
Baseline	33.22	(0.91)	14.17	(0.67)	44.25	(0.75)		
High school travel time not excluded	33.30	(0.93)	14.32	(0.64)	44.23	(0.81)		
+ Measurements added	34.72	(0.95)	14.53	(0.70)	43.72	(1.23)		
			Change in % points				*Change in $1000*	
Repeat policy								
Baseline	9.48	(0.57)	3.94	(0.33)	-1.70	(0.22)	-2.14	(0.26)
High school travel time not excluded	9.73	(0.45)	3.81	(0.33)	-1.75	(0.21)	-2.06	(0.25)
+ Measurements added	10.21	(0.49)	3.89	(0.37)	-1.84	(0.27)	-1.99	(0.21)
Downgrade policy								
Baseline	-9.82	(0.55)	-1.61	(0.25)	-0.30	(0.18)	-1.02	(0.14)
High school travel time not excluded	-10.85	(0.63)	-1.80	(0.23)	-0.39	(0.17)	-1.03	(0.15)
+ Measurements added	-11.04	(0.69)	-1.88	(0.28)	-0.54	(0.21)	-0.95	(0.14)

Note: Predictions from the dynamic model under alternative specifications. Travel times to high school programs added in equations that predict higher education enrollment and graduation in both alternative specifications. The final specification also adds measurements, summarized in Table F4, to the first stage of the estimation procedure. Bootstrap standard errors in parentheses.

the same sign but it is much smaller (and not statistically significant for the work condition). This can be explained by the inclusion of controls for cognitive ability and SES.[42]

Finally, Table F7 shows that results are robust for using two commonly used discount factors in the literature.

[42] A potential downside of adding measurements is that it requires the finite number of types to explain measures that are not of direct importance to the model. I therefore do not use them in the main specification. I also do not use the measures as control variables as they are likely measured with error, not available for everyone and they would make the state-space larger, making it more difficult to obtain good estimates of the CCPs and state transitions in stage 1 of the estimator.

Table F4: Measurements: summary statistics

Variable	Description	Obs	Mean	SD	Min	Max
	Question asked to teacher elementary school (scale of 1 to 5 with 1 the lowest)					
CONS1	Could pay attention in class; has sufficient intellectual capabilities to follow; is smart	3,938	3.64	1.29	1	5
CONS2	Was motivated for school work; wanted to do it really well; worked without reluctance	3,938	3.73	1.24	1	5
CONS3	Could tell a coherent story; explore a topic; stay on the subject	3,936	3.67	1.19	1	5
AGREE1	Did not disturb class intensionally; did not aim to boycott learning	3,917	4.26	1.11	1	5
AGREE2	Held herself to the class rules; waited for her turn; it was not necessary to constantly call her to order	3,935	4.09	1.12	1	5
AGREE3	Was averse to hostilities; was friendly and kind to others; experienced no pleasure in teasing and bullying of others	3,934	3.99	1.10	1	5
EXTRA1	Was open to the teacher; was spontaneous; not defensive	3,935	3.85	1.15	1	5
EXTRA2	Made an energetic and vital impression; looked happy	3,933	3.90	1.06	1	5
EXTRA3	Made contact with fellow students; was open and approachable	3,924	3.99	1.02	1	5
	Definition					
IQ	IQ score, discretized using cutoffs 80, 90, 100, 110 and 120	5,084	3.65	1.36	1	6
Income	Monthly household income in BEF after taxes (1 EUR ≈40 BEF), discretized using cutoffs 40k, 60k, 80k, 100k	5,158	2.47	1.35	1	5
Work	At least one parent is active in the labor market.	4,749	0.86	0.35	0	1

Note: description and summary statistics of measurements of initial traits used in sensitivity checks.

Table F5: Measurements: part 1 of 2

	CONS1		CONS2		CONS3		AGREE1		AGREE2		AGREE3	
Male	0.07	(0.06)	-0.67	(0.07)	-0.23	(0.06)	-1.05	(0.08)	-1.02	(0.07)	-0.82	(0.06)
Language ability	1.62	(0.09)	1.06	(0.07)	1.30	(0.07)	0.37	(0.05)	0.36	(0.05)	0.29	(0.05)
Math ability	0.79	(0.08)	0.58	(0.06)	0.44	(0.06)	0.25	(0.05)	0.29	(0.05)	0.19	(0.05)
High SES	0.48	(0.07)	0.42	(0.08)	0.45	(0.08)	0.08	(0.09)	-0.03	(0.09)	0.03	(0.08)
Type 2	2.24	(0.07)	3.00	(0.08)	2.50	(0.07)	1.90	(0.07)	2.10	(0.07)	2.07	(0.07)
Cut point outcome 2	-3.23	(0.09)	-3.16	(0.09)	-3.33	(0.09)	-3.48	(0.11)	-3.57	(0.10)	-3.58	(0.12)
Cut point outcome 3	-1.01	(0.06)	-1.31	(0.07)	-1.22	(0.06)	-2.26	(0.08)	-2.00	(0.07)	-1.84	(0.07)
Cut point outcome 4	0.62	(0.06)	0.31	(0.06)	0.47	(0.06)	-1.43	(0.07)	-1.05	(0.07)	-0.59	(0.06)
Cut point outcome 5	2.55	(0.07)	2.29	(0.07)	2.65	(0.07)	-0.10	(0.06)	0.59	(0.06)	0.99	(0.06)

Note: Estimates of ordered logit model on measurements used in sensitivity checks. Definition outcome variables: see Table F4. Bootstrap standard errors in parentheses.

Table F6: Measurements: part 2 of 2

	EXTRA1		EXTRA2		EXTRA3		IQ		Income		Work	
Male	-0.55	(0.06)	-0.14	(0.06)	-0.23	(0.06)	0.56	(0.06)	-0.12	(0.05)	0.17	(0.10)
Language ability	0.26	(0.05)	0.31	(0.05)	0.22	(0.05)	1.62	(0.06)	0.14	(0.04)	0.79	(0.08)
Math ability	0.26	(0.05)	0.34	(0.05)	0.21	(0.04)	1.40	(0.08)	0.15	(0.04)	0.09	(0.06)
High SES	0.14	(0.07)	0.32	(0.07)	0.22	(0.07)	0.20	(0.07)	1.66	(0.08)	1.79	(0.25)
Type 2	2.48	(0.07)	2.27	(0.07)	2.02	(0.07)	0.50	(0.06)	0.14	(0.05)	0.16	(0.10)
Cut point outcome 2	-3.10	(0.10)	-3.47	(0.12)	-3.78	(0.14)	-4.02	(0.10)	-0.47	(0.04)		
Cut point outcome 3	-1.35	(0.06)	-1.46	(0.06)	-1.73	(0.07)	-1.98	(0.06)	0.56	(0.04)		
Cut point outcome 4	-0.08	(0.06)	0.06	(0.06)	-0.30	(0.06)	0.14	(0.05)	1.70	(0.05)		
Cut point outcome 5	1.70	(0.06)	1.92	(0.07)	1.49	(0.06)	2.52	(0.06)	2.87	(0.06)		
Cut point outcome 6							5.00	(0.09)				
Constant											1.69	(0.09)

Note: Estimates of (ordered) logit model used in sensitivity checks. Definition outcome variables: see Table F4. Bootstrap standard errors in parentheses.

Table F7: Sensitivity analysis: discount factor

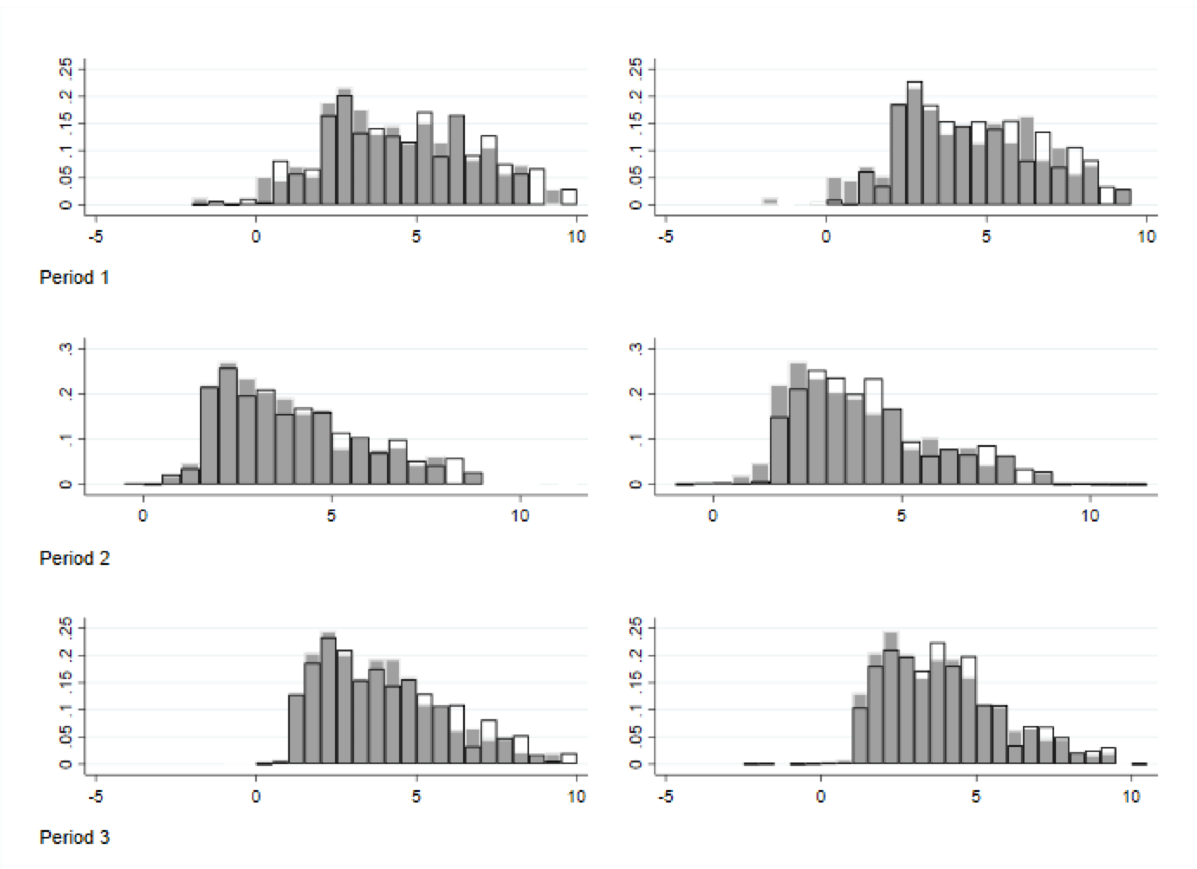
	Study delay		**High school dropout**		**Higher education graduation**		**Student welfare**	
			Predicted value in %					
Status quo								
Baseline ($\beta = 0.9$)	33.22	(0.91)	14.17	(0.67)	44.25	(0.75)		
Alternative ($\beta = 0.95$)	32.95	(0.11)	13.98	(0.68)	44.56	(0.87)		
			Change in % points				*Change in \$1000*	
Repeat policy								
Baseline ($\beta = 0.9$)	9.48	(0.57)	3.94	(0.33)	-1.70	(0.22)	-2.14	(0.26)
Alternative ($\beta = 0.95$)	9.45	(0.53)	3.50	(0.35)	-1.99	(0.21)	-2.53	(0.28)
Downgrade policy								
Baseline ($\beta = 0.9$)	-9.82	(0.55)	-1.61	(0.25)	-0.30	(0.18)	-1.02	(0.14)
Alternative ($\beta = 0.95$)	-10.19	(0.76)	-1.81	(0.28)	-0.26	(0.18)	-1.13	(0.17)

Note: Status quo = students can choose to downgrade or repeat grade after obtaining a B-certificate, Repeat = students must repeat grade after obtaining a B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining a B-certificate. Bootstrap standard errors in parentheses.

G Changes in effort

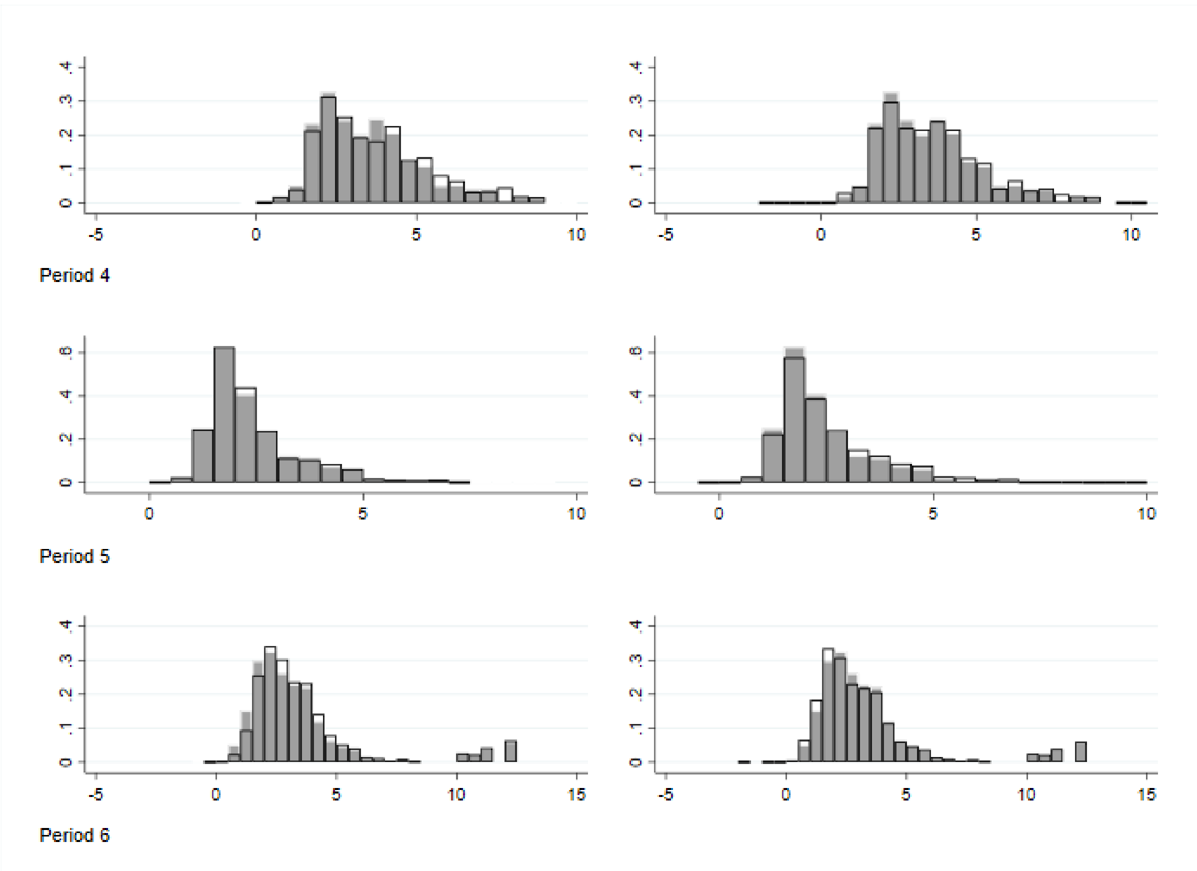
Figures G1, G2 and G3 show the distribution of the log of effort ($\ln(y^*_{it})$) in the counterfactual scenarios. We see a shift to the right in both counterfactuals, especially in early periods. Note however that $\ln(y^*_{it})$ is difficult to compare as students can be in different programs and different grades in the counterfactual scenario, which implies different costs and benefits for the same level of effort. To isolate the incentive effect caused by letting students choose effort, I also show the distribution of the new optimal levels of y^*_{it}, but in the status quo program choices (see Figures G4, G5 and G6). This shows a clearer picture. Without effort choice, we would not see any changes. However, here we clearly see that the new policy gives extra incentives to perform better at the beginning of secondary education, as students are still navigating towards their final track, while there is no impact in later periods.

Figure G1: Changes in effort: period 1 - 3



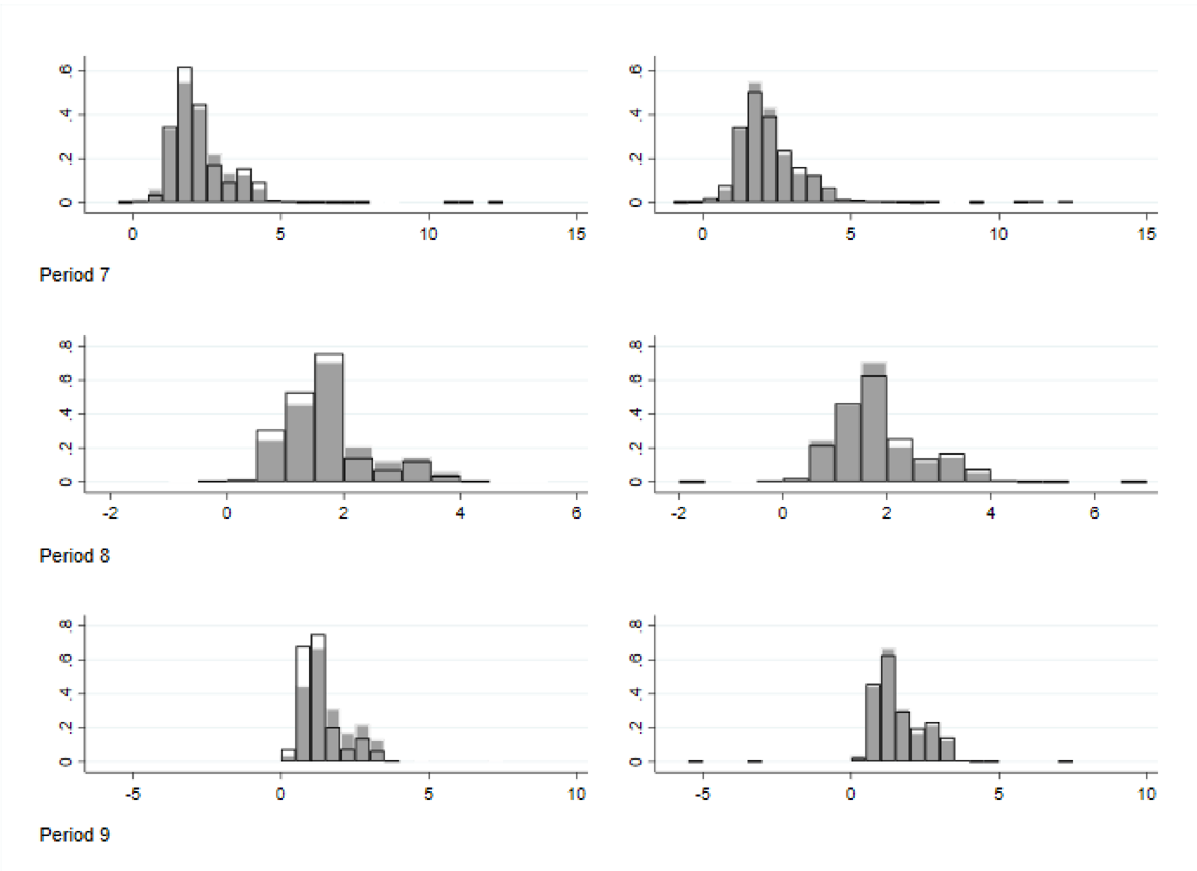
Note: density of the log of effort ($\ln y^*_{it}$) in three scenarios: left = downgrade, right = repeat, gray = status quo.

Figure G2: Changes in effort: period 4 - 6



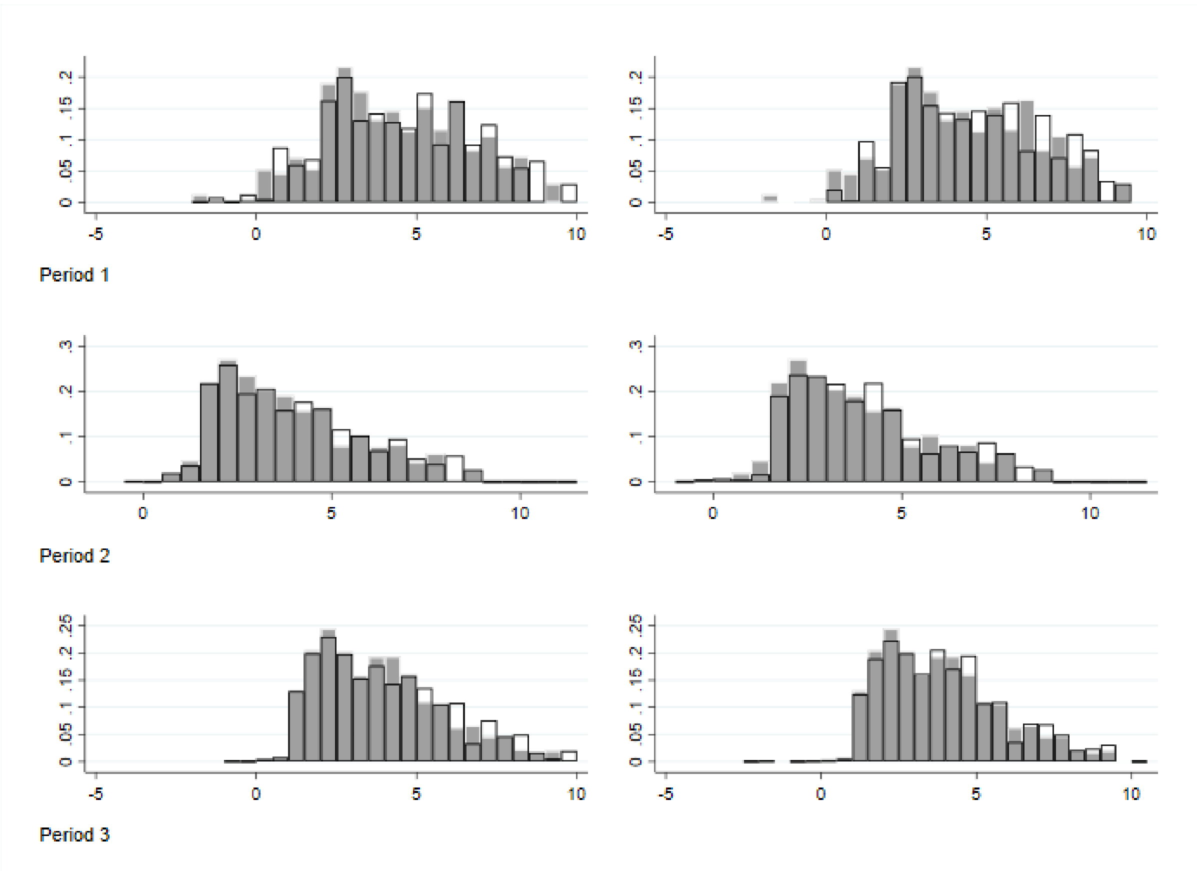
Note: density of the log of effort ($\ln y^*_{it}$) in three scenarios: left = downgrade, right = repeat, gray = status quo.

Figure G3: Changes in effort: period 7 - 9



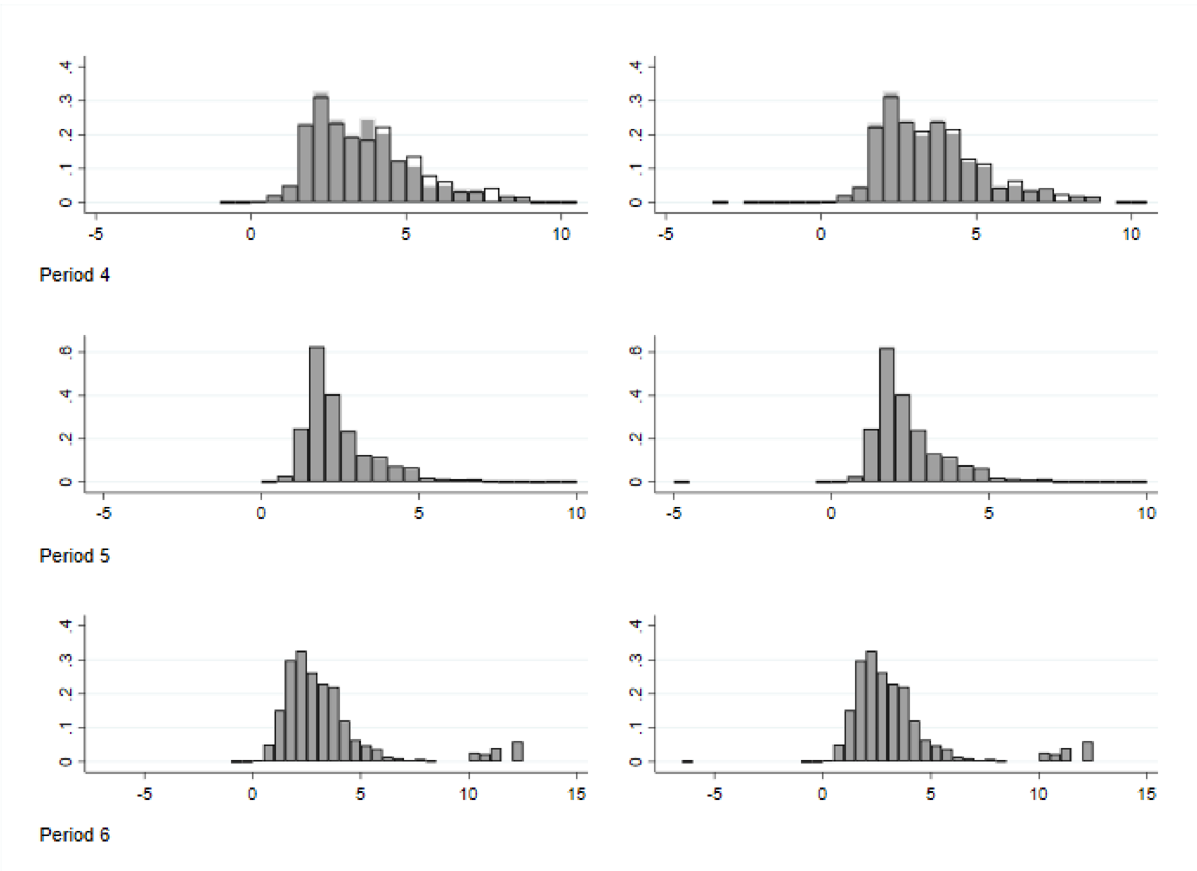
Note: density of the log of effort ($\ln y_{it}^*$) in three scenarios: left = downgrade, right = repeat, gray = status quo. Counterfactual optimal effort levels, but using the status quo predictions of programs and past performance.

Figure G4: Changes in effort for given programs and past performance: period 1 - 3



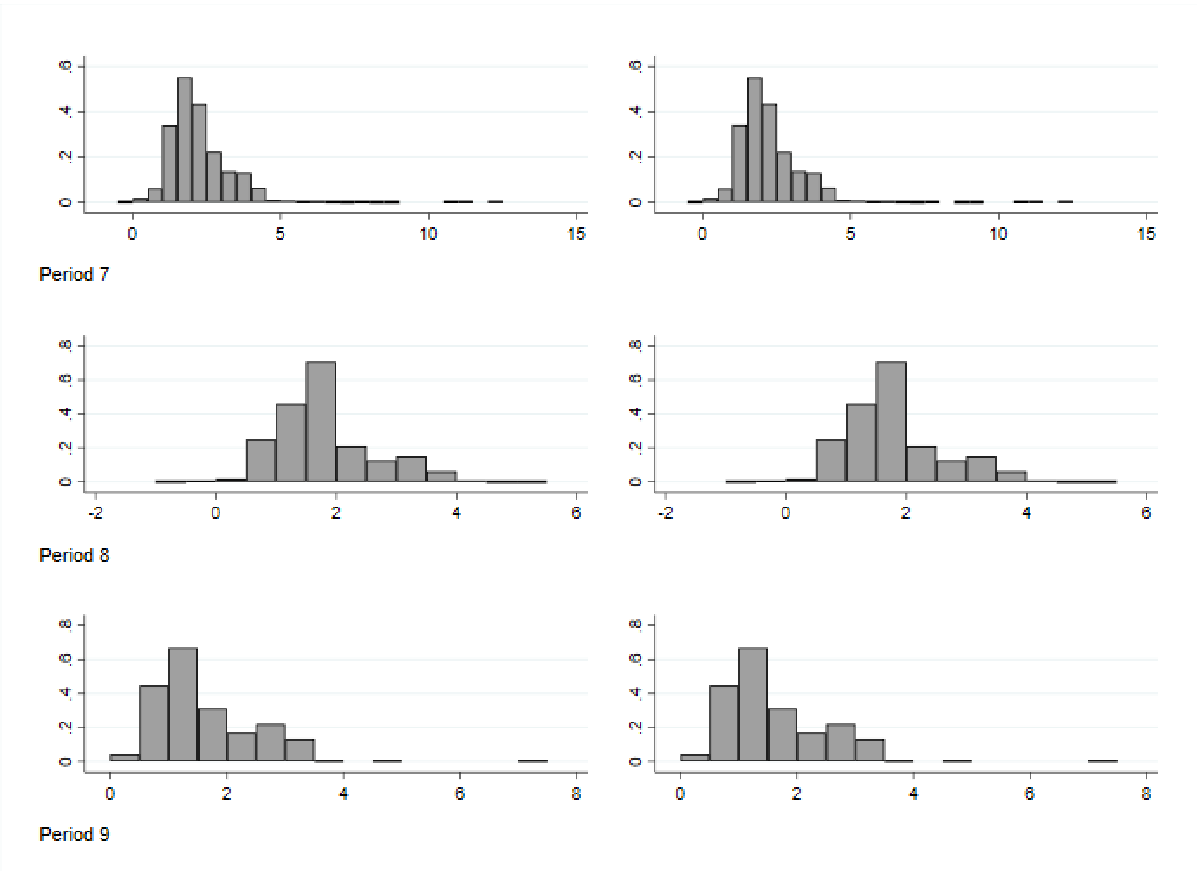
Note: density of the log of effort ($\ln y_{it}^{*}$) in three scenarios: left = downgrade, right = repeat, gray = status quo. Counterfactual optimal effort levels, but using the status quo predictions of programs and past performance.

Figure G5: Changes in effort for given programs and past performance: period 4 - 6



Note: density of the log of effort ($\ln y_{it}^{*}$) in three scenarios: left = downgrade, right = repeat, gray = status quo. Counterfactual optimal effort levels, but using the status quo predictions of programs and past performance.

Figure G6: Changes in effort for given programs and past performance: period 7 - 9



Note: density of the log of effort ($\ln y_{it}^{*}$) in three scenarios: left = downgrade, right = repeat, gray = status quo. Counterfactual optimal effort levels, but using the status quo predictions of programs and past performance.

H SES decomposition

The simulation results of giving low SES students some of the high SES parameters or initial conditions can be found in Table H1. Note that high SES students have higher

Table H1: SES decomposition: differences when low SES have high SES aspects

	Parameters								Characteristics	
	HE		**MC + FC**		**FC**		**MC**		**Observed ability**	
					Change in % points					
SES gap	-7.53	(0.95)	-4.99	(0.60)	-2.46	(0.47)	-2.57	(0.62)	-18.78	(0.68)
Study delay	-0.18	(0.09)	-0.98	(1.06)	2.32	(0.82)	-3.27	(0.64)	-3.34	(0.63)
High school dropout	-0.54	(0.14)	-7.08	(1.26)	-1.36	(1.17)	-5.70	(1.33)	-8.49	(0.52)
HE graduation	5.51	(0.69)	3.65	(0.44)	1.80	(0.34)	1.88	(0.45)	13.65	(0.44)
					Change in $1000					
Student welfare	0.94	(0.16)	9.06	(1.86)	6.39	(1.76)	2.48	(0.44)	4.69	(1.32)

Note: Predictions from the dynamic effort model. Approximated marginal costs are used, based on the specification also used to describe the results in section 5. Decomposition: impact of giving low SES students parameters or variables of high SES students. HE: Higher education enrollment and graduation parameters, MC: marginal cost parameters, FC: fixed cost parameters, observed ability: math and language ability draws from the empirical distribution of high SES. SES gap at graduation: difference in percentage college graduates between high and low SES. Opportunity cost of time: $10/h. Bootstrap standard errors in parentheses.

observed ability. Conditional on this (and other characteristics), they have lower marginal and fixed costs of attending more academic tracks, and are more likely to enroll in and graduate from higher education. In a first simulation, I give low SES students the high SES parameters in higher education. This reflects differences in preferences for obtaining a higher education degree, differences in expected returns to college or differences in support to graduate from it. While college graduation increases, it is still far from high SES students, reducing the gap from 40% to 32%.

In a second set of simulations, low SES students receive the marginal and/or fixed cost parameters of high SES students. The resulting decrease in marginal costs reflects improved circumstances to study. The change in fixed costs reflects different preferences for programs or the influence of parents in the program decision. A combined effect strongly decreases high school dropout and increases higher education graduation by almost 4 percentage points. Note that if only fixed costs decrease, better graduation rates come at the cost of an increase in study delay. This could be a reflection of high SES parents pushing their children into more difficult tracks. If not accompanied by a decrease in marginal cost (i.e. help with studying), it can lead to worse performance.

Finally, we give low SES students the cognitive ability of high SES students. This has the largest impact on outcomes, with an increase in higher education graduation rates of almost 14 % points and reducing the SES gap by almost half.

I Internalities

When students make mistakes, the costs of schooling and the taste shocks cannot be interpreted as utility, but (partly) as optimization mistakes. Consider first the optimal level of effort. If students only cared about utility after high school, they should exert maximum effort to obtain the best performance outcome and graduate without any risk of study delay in the program they want. Marginal costs prevent this. They can be interpreted as a way to make the effort level deviate from what would be optimal if students only cared about the future. Similarly, fixed costs and taste shocks can be interpreted as deviations from the path of program choices students would choose if they only cared about outcomes after leaving high school.

This is easy to see in the simplified model of section 2: students choose to go to high school if $v(x_i, y_i)+\epsilon_{i11} > \epsilon_{i01}$, with $v(x_i, y_i) = u(x_i, y_i)+\beta\gamma+\beta\phi(y_i) \ln (1 + \exp \Psi_1(x_i, g_i = 1))$. If students do not derive any additional (dis)utility from going to school and only care about the future impact of their high school choice, we would have $u(x_i, y_i) = 0$ and $\epsilon_{i11} = \epsilon_{i01} = 0$. Note that $u(x_i, y_i) \equiv -C^0(x_i) - c(x_i)y_i$. $c(x_i) = 0$ would lead to maximum performance. However, it is estimated to match the probability of good performance (and therefore y_i) to what we observe in the data (equation 4). With $c(x_i) = 0$, $C^0(x_i) = 0$ would then mean students do not derive (dis)utility from schooling. However, we need $C^0(x_i) \neq 0$ to match the program choices in the data.[43] Finally, taste shocks $(\epsilon_{i11}, \epsilon_{i01})$ allow for further deviations that cannot be explained by x_i (or, if included, time-invariant types ν_i). What is key for the robustness of counterfactual choice probabilities is that these deviations remain the same in the policies we simulate.

Lavecchia et al. (2016) list four reasons why students make sub-optimal educational decisions: students (1) focus too much on the present, (2) rely too much on routine, (3) care too much about (negative) identities and (4) have too little information about their options. These deviations change our interpretation of estimates

[43] As we do in the application, both marginal and fixed costs can depend on unobserved types too.

but are unlikely to substantially change in the counterfactual policies we consider.[44] Welfare could be substantially different but is expected to be more favorable. First, the main loss in student welfare is coming from taste shocks during high school. If they partly capture mistakes, we should not take them fully into account for the welfare analysis. Second, the expected payoff after leaving high school increases. This is a component that should be weighted more heavily against the other components if students focus too much on the present.[45] Finally, the perceived cost of downgrading, as well as differences between tracks, are likely overvalued by students due to the heavy focus on routine and identity.

[44]More specifically: (1) will make the estimated utility during high school relatively more important than the the payoff after leaving high school. (2) likely causes a high fixed cost of downgrading as it introduces students to a track they are less familiar with. (3) could explain some of the heterogeneity we find in the marginal costs estimates as good/bad performance might be inconsistent with their identity. Similarly identity is likely related to preferences (and therefore the fixed cost estimates) of different tracks. (4) is expected to make high SES students unaware of the higher education options in which they have a good chance to be succesful after leaving from a low track, while low SES students could have the opposite problem. This influences mainly the fixed costs differences between tracks, but also the incentive to perform well and therefore marginal costs.

[45]Since college enrollment goes down but college graduation does not, there can be even further gains for students if we would consider that entering college without graduation is a mistake.

J Tables and Figures

Table J1: Variation in travel time

	Grade													
	7		8		9		10		11		12		13	
Study program	Mean	St.dev	Mean	St.dev	Mean	St.dev	Mean	St.dev	Mean	St.dev	Mean	St.dev	Mean	St.dev
Closest school	31.2	(35.1)	31.6	(35.9)	31.7	(35.4)	30.1	(31.7)	31.4	(32.7)	31.6	(32.6)	41.5	(35.8)
Difference in travel time to closest school with program available														
Academic track														
clas+math					11.1	(16.6)	12.9	(20.6)	14.9	(21.4)	14.5	(21.4)		
clas	4.6	(9.2)	5.1	(9.1)	38.0	(39.7)	39.3	(41.4)	13.9	(19.8)	13.7	(19.8)		
math					4.3	(7.9)	6.5	(14.4)	7.7	(14.8)	7.5	(14.7)		
other	8.2	(13.3)	2.5	(5.6)	8.6	(16.4)	10.2	(18.9)	5.4	(11.9)	5.3	(11.7)		
Middle-Theoretical track														
math					17.4	(18.7)	17.9	(19.4)	18.2	(19.8)	18.0	(19.8)		
other	2.1	(5.7)	8.4	(15.1)	7.9	(13.0)	9.4	(16.2)	8.2	(15.1)	7.8	(14.6)		
Middle-Practical track	9.0	(12.5)	5.4	(9.6)	8.3	(11.2)	11	(16.5)	6.5	(14.6)	6.4	(14.5)		
Vocational track	2.3	(6.9)	1.4	(3.6)	2.0	(5.7)	3.4	(10.2)	2.2	(7.2)	2.2	(7.2)	0	(0)

Note: travel time in minutes per day.

Table J2: High school program and travel time

Study program	**Difference in daily travel time to closest school** Graduates	Others
Academic track	5.1	5.3
clas+math*	9.5	14.8
clas*	11.4	13.9
math	7.6	7.4
other	5.3	5.2
*Middle-Theoretical track**	4.9	6.4
math*	6.5	8.0
other*	6.2	8.0
Middle-Practical track	6.1	6.4
Vocational track	2.3	2.2

Note: travel time in minutes per day. Comparing graduates of the different study programs to other students. * denotes significant difference of means at 5% level. Travel times of grade 12 used.

Table J3: High school program and higher education outcomes: summary statistics

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

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

Table J4: High school program and level and major college degree: summary statistics

	Academic level higher education			Major
	University	Academic college	Professional college	STEM
Study program				
All	*12.4*	*6.0*	*25.5*	*17.8*
Academic				
clas+math	67.0	14.2	13.0	54.0
clas	48.6	10.5	31.4	22.2
math	33.7	18.3	36.2	47.0
other	9.5	6.9	57.8	16.2
Middle-Theoretical				
math	7.2	20.8	44.8	56.0
other	0.7	3.3	43.9	17.5
Middle-Practical	*0.2*	*2.9*	*24.4*	*12.3*
Vocational (13th grade)	*0*	*0.2*	*2.5*	*0.3*

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

Table J5: Higher education program and distance

		Distance: difference with closest HE program			
		Enroll		No enroll	
HE program		Mean	St.dev	Mean	St.dev
University		8.7	(5.8)	8.5	(5.7)
	Antwerp*	41.4	(21.6)	51.7	(21.1)
	Brussels	52.6	(18.5)	51.6	(19.7)
	Ghent*	32.2	(35.5)	86.1	(37.1)
	Hasselt*	8.0	(10.0)	25.0	(32.1)
	Leuven	38.2	(12.8)	37.3	(11.1)
Academic college		4.6	(6.0)	4.1	(5.9)
	STEM	7.1	(7.3)	7.3	(7.4)
	No STEM*	6.2	(6.6)	4.7	(6.3)
Professional college		1.4	(2.9)	1.5	(3.0)
	STEM	3.3	(4.7)	3.5	(4.8)
	No STEM	1.5	(2.9)	1.5	(2.9)

Note: distance in kilometers. HE=higher education. Comparing students that enroll in a campus (for universities) or major (for academic and professional colleges) to students that do not enroll there. * denote significant difference of means at 5% level.

Table J6: Certificates in each track and grade

Track		Grade 7	8	9	10	11	12	13
Academic track								
	A	0.91	0.86	0.92	0.87	0.91	0.97	
	B	0.09	0.12	0.04	0.09	0	0	
	C	0.00	0.02	0.04	0.04	0.09	0.03	
Middle-theoretical track								
	A	0.72	0.67	0.85	0.82	0.85	0.93	
	B	0.28	0.30	0.09	0.11	0.00	0.00	
	C	0.00	0.03	0.06	0.06	0.15	0.07	
Middle-practical track								
	A	0.61	0.56	0.79	0.77	0.81	0.90	
	B	0.39	0.38	0.12	0.14	0.00	0.00	
	C	0.00	0.06	0.09	0.10	0.19	0.10	
Vocational track								
	A	1.00	0.89	0.85	0.88	0.85	0.90	0.85
	B	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	C	0.00	0.11	0.15	0.12	0.15	0.11	0.15

Note: Proportion of students with A-, B- and C-certificates in each track and grade. C-certificate: repeat grade, i.e. all tracks excluded, B-certificate can exclude entire tracks or only elective courses. Note that C-certificates in grade 7 are corrected to be B-certificates that allow the vocational track in grade 8 as this does not require successful completion of the first grade.

Table J7: Exclusions because of certificates (in % of certificates)

	Tracks excluded				**Only elective courses excluded**
Current track	Academic	+Middle-Theoretical	+Middle-Practical	+Vocational	
Academic					
grade 7+8	8.9	4.1	4.0	0.8	2.4
grade 9+10	7.9	4.4	4.4	3.8	2.5
grade 11+12	6.2	6.2	6.2	6.2	0
Middle-Theoretical					
grade 7+8	29.2	21.5	19.7	1.2	0.5
grade 9+10	100	16.5	12.4	6.4	0
grade 11+12	100	11.2	11.2	11.2	0
Middle-Practical					
grade 7+8	41.3	33.8	30.5	3.7	0.5
grade 9+10	100	100	22.3	9.3	0
grade 11+12	100	100	15.1	15.1	0
Vocational					
grade 7+8	100	100	100	7.1	0
grade 9+10	100	100	100	13.8	0
grade 11+12+13	100	100	100	13.6	0

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

Table J8: Impact performance during secondary education

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

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

Table J9: Transitions in educational system (in % of students)

Panel A: transitions in high school

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

Panel B: transitions after high school

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

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

Table J10: Reduced form prediction $\ln y_j^*(x_{it}, \nu_i)$

		Effort in data	
Individual characteristics (baseline = vocational track)			
	Male	-0.324	(0.130)
	High SES	0.592	(0.237)
	Language ability	0.827	(0.112)
	Math ability	0.403	(0.104)
	Type 2	-2.074	(0.323)
Lagged choices			
	Lagged clas	0.205	(0.102)
	Lagged math	0.132	(0.122)
	Lagged track level	0.024	(0.086)
	Downgrade	-0.061	(0.134)
	Upgrade	0.031	(0.155)
Study delay		-0.292	(0.168)
	x track level	-0.044	(0.037)
Repeat		0.568	(0.231)
	x track level	0.276	(0.070)
Travel time high school		0.001	(0.001)
Distance to college			
	Distance to prof college, no STEM	0.014	(0.010)
	Distance to prof college, STEM	-0.022	(0.012)
	Distance to acad college, no STEM	0.030	(0.014)
	Distance to acad college, STEM	-0.005	(0.007)
	Distance to university	-0.005	(0.019)
	Track level x distance to prof college, no STEM	-0.001	(0.005)
	Track level x distance to prof college, STEM	0.003	(0.006)
	Track level x distance to acad college, no STEM	-0.024	(0.007)
	Track level x distance to acad college, STEM	0.004	(0.004)
	Track level x distance to university	0.015	(0.009)
Constants: program (=track + elective)			
Interactions with grade: track, elective			
Interactions with individual characteristics: track, elective, grade			

Note: Estimates of a sample of 5,158 students or 33,239 student-year observations. Grade variable starts counting in high school. Track level = academic level of high school track (0: vocational, 1: middle-practical, 2: middle-theoretical, 3: academic). Bootstrap standard errors in parentheses.

Table J11: Costs of schooling: time, grade and track

	Fixed costs		**Log of marginal costs**	
Time	1	(.)	-0.00	(0.00)
Grade	9.0	(7.8)	0.28	(0.06)
Academic track				
clas+math	76.9	(53.5)	-4.85	(0.84)
clas	-141.8	(47.7)	-3.32	(0.44)
math	-12.6	(49.8)	-3.27	(0.37)
other	-225.3	(49.5)	-1.96	(0.28)
x grade	-22.4	(7.1)	0.07	(0.09)
Middle-theoretical track				
math	102.1	(51.0)	-3.40	(0.47)
other	-181.5	(44.3)	-1.92	(0.25)
x grade	-7.7	(5.0)	-0.12	(0.08)
Middle-practical track	-25.4	(43.9)	-1.65	(0.34)
x grade	-21.5	(4.9)	-0.22	(0.08)
Vocational track	112.6	(46.3)	-4.25	(0.53)
Part-time	270.6	(31.7)		

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

Table J12: Costs of schooling: student characteristics and elective courses

	Fixed costs				**Log of marginal costs**			
	Interaction with classical languages		Interaction with intensive math		Interaction with classical languages		Interaction with intensive math	
Male	-4.1	(8.3)	-50.8	(11.7)	0.53	(0.36)	0.87	(0.27)
Language ability	-57.1	(11.1)	29.6	(13.2)	-1.11	(0.41)	-0.82	(0.34)
Math ability	-21.8	(9.1)	-71.0	(14.9)	-0.48	(0.40)	0.54	(0.32)
High SES	-36.2	(9.5)	-25.4	(11.3)	-0.48	(0.39)	0.26	(0.30)
Type 2	83.5	(12.5)	35.5	(11.8)	-0.19	(0.43)	-0.02	(0.33)

Note: Estimates of a sample of 5,158 students or 33,239 student-year observations. Scale = minutes of daily travel time. The marginal costs in the model are a flexible function of state variables, this table summarizes them by regressing their logarithmic transformation on the same variables that enter the fixed costs. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Clas= classical languages included. Math= intensive math. Bootstrap standard errors in parentheses.

Table J13: Type probabilities in %

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

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

Table J14: Performance thresholds track

	Threshold	
Increase to obtain outcome 3	0.871	(0.036)
Increase to obtain outcome 4	1.102	(0.043)
Increase to obtain outcome 5	1.744	(0.054)

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

Table J15: Performance elective courses

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

Note: Bootstrap standard errors in parentheses.

Table J16: Value of obtaining degree

	Degree values	
High school degree	512.2	(101.3)
x level	108.1	(60.7)
x vocational	-146.8	(99.3)
12th grade certificate vocational track	529.8	(65.5)

Note: Estimates of μ^{degree}. Scale = minutes of daily travel time. Level = academic level of high school track (0: vocational, 1: middle-practical, 2: middle-theoretical, 3: academic). Bootstrap standard errors in parentheses.

Table J17: Estimation results higher education (1)

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

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

Table J18: Estimation results higher education (2)

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

Note: Estimates of higher education outcomes as specified in Appendix D. Clas= classical languages included. Math= intensive math. High school level = academic level of high school track (0: vocational, 1: middle-practical, 2: middle-theoretical, 3: academic). HE Level = level of higher education (average ability). Bootstrap standard errors in parentheses.

Table J19: Estimation results higher education (3)

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

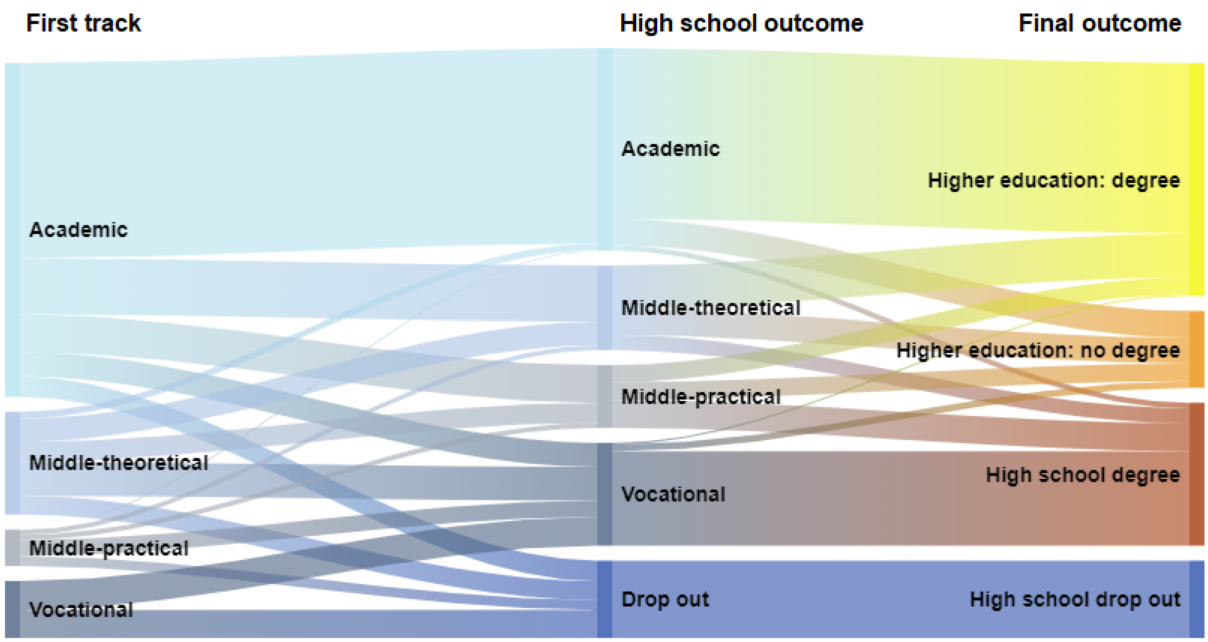
Note: Estimates of higher education outcomes as specified in Appendix D. HE Level = level of higher education (average ability). Bootstrap standard errors in parentheses.

Table J20: OLS regressions on initial conditions and counterfactuals

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

Note: OLS regression of predictions from the dynamic model. B-certificate = students acquired skills to proceed to next grade but only if they downgrade, i.e. switch to track of lower academic level or drop an elective course. Status quo = students can choose to downgrade or repeat grade after obtaining B-certificate, Repeat = students must repeat grade after obtaining B-certificate, Downgrade = students must downgrade and not repeat grade after obtaining B-certificate. Ability measured in standard deviations. Type 2 = dummy equal to one if student belongs to unobserved type 2 instead of 1. High SES= at least one parent has higher education degree. Opportunity cost of time: $10/h. Bootstrap standard errors in parentheses.

Figure J1: Transitions in the educational system



Note: Left: program chosen at the start of secondary education, middle: option in which student graduated or drop out, right: highest educational outcome. Size of the flows proportional to number of students transitioning. See Appendix Table J9 for the corresponding data. Students in the vocational track can attend a 13th grade but are considered graduates in this figure after successfully completing grade 12. As in Declercq and Verboven (2018), I define a higher education degree as three successful years of higher education in a time span of six years. Figure created using Google Charts.