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.