On Track Mobility, Grade Retention and Secondary School Completion*

Stijn Baert[†] Bart Cockx[‡] Matteo Picchio[§]

October 4, 2013

Abstract

In this article we model schooling attainment and track choices to understand the role played by past outcomes on subsequent performances in secondary school. We use data on 4,214 students in Flanders (Belgium). We estimate dynamic qualitative choice models which flexibly take into account the presence of unobserved characteristics jointly determining the educational choices and performances. In contrast to most of previous findings, we find that grade retention has a positive impact on the next evaluation and can permanently affect subsequent educational achievements. The direction of the permanent effect is essentially heterogeneous: while more able students are permanently penalised by retention, less able students benefit from it. We conclude that in the design of the optimal retention policy, the interaction between retention and students' abilities should be taken into account.

Keywords: Education, track mobility, grade retention, dynamic qualitative choice models, heterogeneous treatment effect.
JEL classification codes: C33, C35, I21.

*Matteo Picchio acknowledges financial support by the Research Foundation - Flanders (FWO) when he was at Ghent University from October 2011 until October 2012. We also wish to thank the participants to the Labour Health seminar in Tilburg (April 2013), to the seminar in Sevilla (April 2013), to the DSSE-ISFOL workshop on Human Capital and Education in Rome (July 2013), to the CESifo conference in the Economics of Education in Munich (September 2013), to the EALE Conference in Turin (September 2013), and to the AIEL conference in Rome (September 2013).

[†]Sherppa, Ghent University. E-mail: Stijn.Baert@UGent.be.

[‡]Sherppa, Ghent University; UCLouvain (IRES), Louvain-la-Neuve; IZA, Bonn; CESIfo, Munich. E-mail: Bart.Cockx@UGent.be.

[§]Department of Economics and Social Sciences, Marche Polytechnic University; CentER, Tilburg University; Sherppa, Ghent University; IZA, Bonn. E-mail: m.picchio@univpm.it.

1 Introduction

Two of the most notable differences between school systems across OECD countries consist in grade retention policies and in how pupils are separated into different curriculum tracks.[1] Grade retention is used in many countries as a tool to improve poor academic performances. The hypothesis is that, by resitting the same grade, low-achieving students have extra time to catch up to the grade-level requirements, in terms both of knowledge and emotional maturity. By having more time to develop the skills needed in the subsequent grades, resitting students should be less at risk of failure in the future. Moreover, the threat of retention might be an incentive device to work more diligently and harder. However, retention might generate personal and academic costs with both short- and long-term effects, since it might: hurt pupils' self-esteem (Browman, 2005; Byrd et al., 1997); generate psychological costs of separating students from their peers (Alexander et al., 1994); produce financial costs to the families and to society in terms of teaching resources (Eide and Goldhaber, 2005).

Tracking according to pupils' abilities and interests is a quite common practice to take into account the pupils' diversity of skills and preferences. Tracking systems are quite heterogeneous across OECD countries in terms both of the number of available educational tracks and of the timing of tracking. In Europe, tracking mostly takes the form of stratified curricula, both within and between schools, specialised either in general or vocational education. It is often argued that tracking, by generating groups that are more homogeneous in terms of abilities and tastes, increases the effectiveness and efficiency of instruction (Hallinan, 1994). However, recent evidence shows that (early) tracking reduces the performance of students and increases the impact of family background on student performance (Aakvik et al., 2010; Guyon et al., 2012; Hall, 2012; Kerr et al., 2013; Meghir and Palme, 2005; Piopiunik, 2013). This evidence is related to several theoretical disadvantages of tracking. First, the opponents of school tracking argue that it exposes students to the risk of ending up into the wrong track, especially in case of hierarchical ordering. In this case allocation to the wrong track is on the one hand irreversible if the initial track choice is too low and on the other hand costly if the initial choice is too high. Second, tracking might increase inequality since pupils from weak social backgrounds choose lower tracks. Third tracking might forgo positive spillovers between weak and strong students. Last, teachers' instructional practices could vary across tracks and generate different schooling achievements (Gamoran, 1989; Gamoran et al., 1995). Van Houtte (2004) finds indeed that in Flanders, the Dutch speaking region of Belgium, the chances of failing are higher in technical and vocational schools than in general schools.

[1] See OECD (2004, p. 262) for a comparison of the features of school systems of OECD countries.

In the present study we use econometric modelling tools and identification analysis to examine the interrelated dynamics of secondary school grade retention, track choices and achievements of a sample of Belgian pupils living in Flanders. The secondary education system of Flanders is characterised by grade retention for low achieving students and hierarchical tracks with only downgrades permitted ('cascade' system). We also shed light on the role played by family background and unobserved abilities, especially looking at how unobserved abilities interact with retention episodes in determining schooling pathways.

The empirical analysis is carried out using the SONAR dataset, a retrospective survey conducted in Flanders on the 1976, 1978 and 1980 cohorts. The SONAR dataset contains very rich information on education, but also on family and labour market experiences. Our sample is made up of 4,214 students belonging to the 1978 and 1980 cohorts. We exploit the ample information on secondary school performances and choices available for these two cohorts to estimate dynamic qualitative choice models.

The identification of the interrelated dynamics between grade retention, track mobility and schooling attainment is obtained by addressing some key challenges. First, educational achievements and choices are likely to be determined by a set of unobserved determinants, e.g. behavioural and cognitive skills, with an unknown correlation structure. In order to disentangle the pure effects of past educational outcomes on future ones from the spurious effects determined by unobserved abilities, we take into account the presence of unobserved heterogeneity by semi-parametric maximum likelihood techniques (Heckman and Singer, 1984; Mroz, 1999). Second, at the start of secondary school pupils have already different years of delay due to retention episodes either in kindergarten or in primary school. If we assume that grade retention affects future outcome variables, we have an initial conditions problem. The years of delay at the beginning of secondary school cannot be easily assumed to be exogenous, since they are very likely correlated to the unobserved determinants. We solve for initial conditions by adding an equation for the years of delay at the beginning of secondary school which depends on unobserved heterogeneity and an exclusion restriction (Heckman, 1981). Third, as pointed out by Fruehwirth et al. (2011), the effect of grade retention might be heterogeneous and vary by students' unobserved abilities. We allow therefore the effect of past retention episodes to vary across different levels of the unobserved determinants. Fourth, there might be sample selection attrition induced by students dropping out of secondary school. We model therefore also the probability of exiting school without the secondary education diploma at the end of each year from the end of compulsory education onwards, where the unobserved components determining the school drop-out is allowed to be correlated to the unobserved determinants of the other endogenous processes.

On the one hand our results show that pupils downgrading track are more likely to get good evaluations in the next academic year. Moreover, students who have experienced a track change are less likely to downgrade in the following year, meaning that downgrading stabilises the track pathways of students. On the other hand, and in contrast to most of previous findings, we find that grade retention has a positive impact on the next evaluation and can permanently affect subsequent performances. The direction of the permanent effect depends on unobserved heterogeneity. While more able students are permanently penalised by retention, less able students benefit from it. We conclude that when looking for the optimal retention policy, the interaction effect between retention and students' abilities should be taken into account.

This study is organised as follows. In Section 2, we present the educational system of secondary school in Flanders (Belgium). Section 3 describes the data and summarises basic descriptive statistics of the variables used in the empirical analysis. Section 4 presents the econometric model. Section 5 reports the estimation results. Section 6 concludes.

2 The Flemish Secondary School Educational System

In this paper we use data from Flanders, the Dutch speaking region of Belgium, situated in the northern part of the country. Belgium is a federal country with several competences devolved to its three Regions (Flanders, Brussels and Wallonia) and three Communities (Flemish, French and German speaking). While the federal authorities are competent for all matters of national importance, territorial and person-related issues are left to Regions and Communities. The Flemish Community is in charge of all aspects of education policy in Flanders.

Nationwide, the Belgian Constitution states that every child has the right to education, which is granted by a compulsory education law. Compulsory education starts on 1 September of the year in which the child turns 6 years old and ends on 30 June of the year in which (s)he reaches the age of 18.[2] Children start primary school in the year in which they turn 6 years old. However, they might start one year earlier or some years later if in kindergarten they are suggested to do so.[3] Grade retention and grade skipping are also allowed in primary school. Hence, pupils may start secondary school at different ages. In case of no retention or skipping in primary school and regular age at the beginning of primary school, pupils start secondary school in the year they turn 12 years old.[4]

[2] Starting from the age of 15 (conditional on passing the first two years of full-time secondary education) or the age of 16 (unconditionally), only part-time education is mandatory.

[3] In our sample, 1.4% of children started primary school in the year they turned 5 and 1.1% started it when 7 or 8.

[4] Out of 4,214 pupils in our sample, only 46 (1.1%) started secondary school in the year they turned 11

In Flanders, when entering secondary school, students formally choose between hierarchical ordered tracks. Students are grouped or tracked according to their abilities and interests, a quite common practice in OECD countries to take into account the diversity of skills and preferences of pupils in education. In this paper, as in Van de gaer et al. (2006) and Van Houtte et al. (2012), we refer to 'tracking' as the situation in which students are taught entirely different curricula depending on their curriculum choice which may be restricted after unsatisfactory performances. This is different from 'setting' or 'banding', where pupils in the same curriculum are taught at different level of difficulties given their ability (Gamoran et al., 1995). The Flemish secondary school system consists of several tracks which can be divided into four main education forms: i) general education (ASO) which emphasises general education and provides firm foundations for tertiary education; ii) technical education (TSO) which provides general foundations for practising a profession; iii) art education; iv) vocational education (BSO) which is oriented to the accumulation of skills for a specific profession. In this study, we do not consider the art education track, because of the small number of pupils in our sample choosing it. Our analysis is limited to track choices and track mobility between ASO, TSO and BSO. Students obtain the secondary school diploma if they successfully pass the 6 grades of ASO and TSO and the 7 grades of BSO. All the secondary school diplomas give access to tertiary education.

Track mobility in secondary school is allowed with the following constraints and features. First, track change is not permitted at the beginning of the last grade, hence at the beginning of grade 6 for the ASO and the TSO tracks and grade 7 for the BSO track. Second, tracks are hierarchical and moving upward is not allowed, i.e. it is not possible to go from BSO to TSO/ASO or from TSO to ASO. It is anyway possible at the beginning of each academic year to downgrade the track and move from ASO to TSO/BSO and from TSO to BSO. Finally, track mobility is also possible at a finer level within the ASO, TSO and BSO tracks. Within each major track, it is indeed possible to identify hierarchical subtracks with different curricula of different complexity for which the just mentioned track mobility constraints are satisfied. The data at hand allow us to identify two hierarchical subtracks for ASO, which we name ASO$+$ and ASO$-$ and two hierarchical subtracks for TSO, labelled TSO$+$ and TSO$-$.

At the end of each academic year, pupils receive an evaluation: A, B, or C. Pupils getting an A can access the next grade and, if they wish, can downgrade the track. Pupils obtaining a C must resit the grade and, if they wish, can downgrade the track. Pupils getting a B can decide whether to resit the grade or not. If they decide to resit, they can stay in the same track. If they decide not to resit the grade, they must downgrade the track.

and 176 (4.2%) started secondary school with delay.

Given the set-up of the Flemish educational system, there are different choices that pupils (or/and their parents) have to make in each academic year. First, they have to decide the track. Second, if at the end of the year they get a B, they have to decide whether to resit the grade or not. Finally, they have to decide whether to downgrade the track. Once they turn 18 years old, they can also choose to drop-out the school without the diploma. We will model all these choices and students' performances (evaluation and secondary school completion) in a multiple-equations dynamic model for categorical outcome variables, where past choices and past performances are allowed to affect future schooling pathways.

3 Data and Sample

The dataset used in the empirical investigation comes from the SONAR survey. The SONAR survey retrospectively collected information on education, family background, family formation and labour market experiences for a sample of almost 9,000 of individuals living in Flanders and born in 1976, 1978, or 1980.[5] The 1976 cohort was interviewed thrice, at age 23, 26 and 29. The 1978 cohort was interviewed twice, at age 23 and 26. Finally, the 1980 cohort was interviewed only once at age 23. While we only know starting and ending years of primary school, for secondary school we have detailed information, year by year, on school track choices, evaluations, school drop-out and obtaining the diploma.

Since there is no detailed information on tracks for the 1976 cohort, we removed it from the sample and are left with 5,953 pupils. In order to have a sample of pupils with a homogeneous educational, social and family background, we removed from the sample pupils whose grandmother on mother's side have a foreign nationality, pupils who need special help, temporarily or permanently, and are therefore in special schools and pupils who start secondary school when older than 15. We also deleted those entering the art curriculum, those reporting a break of one or more years in secondary school attendance, those leaving school before the end of compulsory education and those with inconsistent or missing information on the progression of the grade, evaluation and grade mobility. After applying these selection criteria, we end up with a sample of 4,214 pupils. The exit

[5] A study of the representativeness of the sample was conducted by the SONAR group and reported in SONAR (2000). The sample is representative with respect to gender. Comparing the sample with respect to other characteristics is more difficult because of a lack of comparable data. A cautious comparison with statistics of the Ministry of Education and the Labor Force Study reveals that the sample is representative with respect to family formation. The lower educated, the unemployed and respondents from lower social classes are instead somewhat under-represented.

from secondary school might take place with or without the diploma. In our sample there are students who are retained multiple times; the observed maximum number of years in secondary school is 11. If students move to part-time education, they are censored in the year they move to it. Hence, we use all the information until the transition to part-time education, but we disregard all the information from the moment of entering part-time education.[6]

Table 1 reports summary statistics of schooling attainment and choices which we model in the empirical analysis. First, we report some outcomes and decisions at the end of the schooling year averaged over the secondary education career. In our sample on average almost 90% get an A, the highest evaluation, while about 6% and 4% are assigned a B and a C, respectively. Around 5% of the pupils are retained on average at the end of the academic year.[7] We hardly see track transitions involving downgrades of more than 2 steps: only 48 track transitions involve a downgrade of three steps and nobody makes a 4-step downgrade, i.e. from ASO to BSO. Hence, given the starting track, information on track changes compressed in no downgrade, 1-step downgrade and 2-step downgrade is able to describe almost all the possible track transitions. In the 90% of the cases, pupils stay in the same track, while 7.5% of the students start the new year with a 1-step track downgrade and 2.5% with a 2-step downgrade. Second, Table 1 shows the average cumulative delay at the beginning of grade 1 and grade 2 and at the end of secondary school (irrespective of whether one exits with or without a diploma). At the beginning of secondary education, the average number of years schooling delay is 0.03. No student is retained at this first grade. By the end of secondary school pupils are on average retained for 0.32 years. Third, Table 1 reports the relative frequency of track choices at the beginning of grade 1 and grade 2. At the beginning of grade 1, we have only partial information about the school track choice. We only know whether the student is in the vocational track (BSO) or not (ASO/TSO). This partial observability generates a complication in modelling track choice at the start of grade 1 and subsequent downgrades. We explain how we deal with it in Subsection 4.4. Only starting from grade 2, we have more detailed information on the tracks and we can group track choices into 5 hierarchical categories: ASO+, ASO−, TSO+, TSO− and BSO. At the beginning of grade 1, 6.3% of pupils choose BSO. As a result of some downgrading decisions, this frequency increases almost up to 10% when moving to the second grade; 27% are instead in ASO+, 40% in ASO− and the remaining 23% is split almost evenly between TSO+ and TSO−. By the end of secondary education 19% of the pupils are in BSO, 13% are in

[6]Since only 184 students left full-time education for part-time education, we preferred not to model their transition to part-time education and their future schooling experiences.

[7]This figure is in line with the figures reported in OECD (2004, p. 262) for the whole Belgium.

ASO+, 36% in ASO−, 11% in TSO+ and 22% in TSO−. Finally, out of the 4,214 pupils who start secondary school, 4.4% enter part-time education and are therefore censored in our model, 86.5% are able to get the full-time secondary education diploma, while the remaining 9.2% drop out of secondary school without the diploma.

Table 1: Summary Statistics of the Outcome Variables: Schooling Attainment and Choices

	Mean	Std. Dev.
Outcomes and decisions at the end of the year[§]		
Evaluation: A	0.897	0.304
Evaluation: B	0.059	0.235
Evaluation: C	0.044	0.206
Retention	0.054	0.226
No downgrade	0.900	0.300
1-step downgrade	0.075	0.263
2-step downgrade (or more)	0.025	0.157
Cumulative delay		
Cumulative delay at the beginning of grade 1	0.031	0.228
Cumulative delay at the beginning of grade 2	0.031	0.228
Cumulative delay at the end of secondary education	0.319	0.624
Track at the beginning of grade 1		
ASO/TSO	0.938	0.242
BSO	0.063	0.242
Track at the beginning of grade 2		
ASO+	0.272	0.445
ASO−	0.403	0.490
TSO+	0.095	0.293
TSO−	0.132	0.339
BSO	0.098	0.298
Track at the end of secondary school		
ASO+	0.133	0.340
ASO−	0.360	0.480
TSO+	0.105	0.306
TSO−	0.218	0.413
BSO	0.185	0.388
Exit from secondary school		
With diploma	0.865	0.342
Without diploma	0.092	0.289
Censored to part-time education	0.044	0.204
Number of pupils	4,214	
Number of pupils × no. of years of schooling	26,313	

[§] Yearly averages over the secondary education career.

Table 2 reports descriptive statistics of the covariates used in the econometric analysis. Most of the pupils start primary education in time (97.6%), i.e. in the year they turn 6. The fraction of those starting in time secondary school is smaller and equal to 94.7%, while the fraction of those who start late rises from 1.3% in primary school to 4.2% in secondary due to retention in primary school. Almost one half of the pupils have a sibling, 13.8% are only child and almost 40% have more than one sibling. Pupils' fathers are more educated than pupils' mothers, having in average 6.2 years of successful education beyond primary

Table 2: Summary Statistics of Covariates at the Beginning of Secondary School

	Mean	Std. Dev.
Female	0.502	0.500
Calendar day of birth	183.9	104.8
Father's education (years)[§]	6.201	3.339
Mother's education (years)[§]	5.809	3.032
Age at the beginning of primary school		
5 years old	0.014	0.115
6 years old	0.976	0.154
7 years old	0.011	0.103
8 years old	0.0002	0.014
Age at the beginning of secondary school		
11 years old	0.011	0.104
12 years old	0.947	0.223
13 years old	0.042	0.200
Cohort		
1978	0.497	0.500
1980	0.503	0.500
Presence of siblings		
0	0.138	0.345
1	0.465	0.499
2	0.257	0.437
3 or more	0.140	0.347
Number of pupils	4,214	

[§] Father's and mother's education measures the number of successful schooling years beyond primary school, which lasts 6 years.

school against 5.8 years for mothers.

4 The Econometric Model

In this section, we write down the likelihood function and clarify the identifying assumptions. Finally, we deal with the problem of partial observability of tracks at the start of secondary school.

4.1 Model Specification and the Likelihood Function

If we aim at understanding the determinants of educational achievements in secondary school, we have to take into account that many determinants are potentially endogenous variables. For example, the total number of years of delay with which students start each grade and the different track choices they make might influence future schooling attainment and decisions, but are at the same time the results of past performances and choices. Performances and choices might be correlated across equations and over time due to the presence of unobserved heterogeneity. If we wish to disentangle the causal effects from the spurious ones, we have to control for it.

The six outcome variables that we model for each student i at each academic year t, with $i = 1, \cdots, N$ and $t = 1, \cdots, T$ are:

- Track choice at the beginning of secondary school (tr_i). Since tracks are hierarchically ordered, tr_i is an ordered response taking on the increasing values $\{\text{BSO}, \text{TSO}-, \text{TSO}+, \text{ASO}-, \text{ASO}+\}$.
- Evaluation at the end of each academic year (ev_{it}) or, if in the last grade, the success in getting the diploma (di_{it}). ev_{it} is an ordered response taking on the increasing values $\{\text{C}, \text{B}, \text{A}\}$. di_{it} is instead binary and equal to 1 if the student gets the diploma at the end of the academic year or equal to 0 if (s)he fails the last grade and has to resit.
- School drop out (out_{it}) if turning 18 (compulsory schooling age) in that calendar year or older than 18. out_{it} is a dummy indicator equal to 1 if the student drops out of school, 0 otherwise.
- Resitting decision (re_{it}) if the evaluation is B ($ev_{it} = \text{B}$). re_{it} is a dichotomous variable equal to 1 if the student chooses to resit when (s)he gets a B, 0 if (s)he chooses instead to downgrade.
- Track downgrade (dow_{it}) which is defined as an ordered response taking values on $\{0, 1, 2\}$, where 0 means 'no downgrade', 1 stands for '1-step downgrade' and 2 is '2-step downgrade'.

Furthermore, we have an initial conditions equation for the number of years of delay (in_i) at the beginning of secondary school. As mentioned before, pupils start secondary school at different ages due to different past retention histories either in primary school or in kindergarten. If we assume that past performances like past grade retention affect future outcome variables, we have an initial conditions problem. The years of delay at the beginning of secondary school cannot be easily assumed to be exogenous, since they are very likely correlated to the unobserved determinants. We solve for initial conditions by adding an equation for the years of delay at the beginning of secondary school which depends on unobserved heterogeneity and an exclusion restriction (Heckman, 1981). in_i takes values on $\{-1, 0, 1\}$. It is equal to 0 when the student starts secondary school without delay, i.e. in the year in which (s)he turns 12, to -1 if one year in advance and to 1 if one year late.

Let $\mathbf{Y}_{it} \equiv (ev_{it}, di_{it}, out_{it}, re_{it}, dow_{it})$ be the vector collecting the five time-varying outcome variables and $\mathbf{z}_i$ be the vector of observed explanatory variables. Denote by $\mathbf{v}_i \in \mathbb{R}^7$ a random vector of equation-specific time-invariant covariates that are unobserved to the analyst. This vector of unobserved determinants has an unknown cumulative distribution function G.

We can always write the density of $(in_i, tr_i, \mathbf{Y}_i)$ conditional on $(\mathbf{z}_i, \mathbf{v}_i)$ as

$$\begin{aligned} f(in_i, tr_i, \mathbf{Y}_i|\mathbf{z}_i, \mathbf{v}_i) &= f(in_i|\mathbf{z}_i, \mathbf{v}_i) \cdot f(tr_i|\mathbf{z}_i, \mathbf{v}_i, in_i) \\ &\quad \cdot \prod_{t=1}^{T} f(\mathbf{Y}_{it}|\mathbf{z}_i, \mathbf{v}_i, \mathbf{Y}_{it-1}, \cdots, \mathbf{Y}_{i1}, tr_i, in_i) \\ &= f(in_i|\mathbf{z}_i, \mathbf{v}_i) \cdot f(tr_i|\mathbf{z}_i, \mathbf{v}_i, in_i) \cdot \prod_{t=1}^{T} f(\mathbf{Y}_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}), \quad (1) \end{aligned}$$

where $\Im_{it-1}$ denotes the information set containing all the realisations of the endogenous variables from $t-1$ until the beginning of the processes, i.e. $\Im_{it-1} = (\mathbf{Y}_{it-1}, \cdots, \mathbf{Y}_{i1}, tr_i, in_i)$.

Assumption 1 *(Sequentiality):*
Within each academic year t and for $t = 1, 2, \cdots, T$, the five time-varying outcome variables in $\mathbf{Y}_t$ are realised sequentially with the following chronological order: performance at the end of the year, either evaluation or achieving the diploma, ($ev_t \vee di_t$); school drop-out decision (out_t); resitting decision (re_t); track downgrade decision (dow_t).

Given Assumption 1 on the sequentiality of the realisations of the endogenous variables, it is meaningful to rewrite the conditional density in Equation (1) as

$$\begin{aligned} f(in_i, tr_i, \mathbf{Y}_i|\mathbf{z}_i, \mathbf{v}_i) &= f(in_i|\mathbf{z}_i, \mathbf{v}_i) \cdot f(tr_i|\mathbf{z}_i, \mathbf{v}_i, in_i) \\ &\quad \cdot \prod_{t=1}^{T} \Big[f(ev_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1})^{1-g_{it}} f(di_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1})^{g_{it}} \\ &\quad \cdot f(out_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it})^{s_{it}} \\ &\quad \cdot f(re_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it} = B)^{1-g_{it}} \\ &\quad \cdot f(dow_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it}, re_{it})^{c_{it}} \Big], \quad (2) \end{aligned}$$

where g_{it} is an indicator variable equal to 1 if the student is in the last grade of secondary school and 0 otherwise, s_{it} is a dummy equal to 1 if the student belongs to the set at risk of school drop-out (legally allowed to drop out) and c_{it} is equal to 1 if the student is in the ASO/TSO tracks and 0 if (s)he is in the BSO track (BSO students do not have the option to downgrade as already at the bottom of the track hierarchy).

We cannot derive the likelihood function on the basis of the density in Equation (2), because we do not observe $\mathbf{v}_i$. Instead, we integrate $\mathbf{v}_i$ out after assuming that it is orthogonal to $\mathbf{z}_i$.

Assumption 2 *(Orthogonality):*
$\mathbf{v}_i \perp\!\!\!\perp \mathbf{z}_i$.

Under Assumption 2 on the orthogonality between the exogenous covariates and the unobservables we can integrate $\mathbf{v}_i$ out once we specify its cumulative distribution function $G(\mathbf{v}_i)$, yielding the following marginal density:

$$
\begin{aligned}
f(in_i, tr_i, \mathbf{Y}_i|\mathbf{z}_i) = \int_{\mathbb{R}^7} & f(in_i|\mathbf{z}_i, \mathbf{v}_i) \cdot f(tr_i|\mathbf{z}_i, \mathbf{v}_i, in_i) \\
\cdot & \prod_{t=1}^{T} \Big[f(ev_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1})^{1-g_{it}} f(di_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1})^{g_{it}} \\
\cdot & f(out_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it})^{s_{it}} \\
\cdot & f(re_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it} = B)^{1-g_{it}} \\
\cdot & f(dow_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it}, re_{it})^{c_{it}} \Big] \mathrm{d}G(\mathbf{v}_i). \qquad (3)
\end{aligned}
$$

Providing an empirical specification to the each of the probability density functions in Equation (3) leads to the sample log-likelihood function:

$$
\begin{aligned}
\ell(\boldsymbol{\theta}, \boldsymbol{\delta}) &= \sum_{i=1}^{N} \ln \Big[\int_{\mathbb{R}^7} \mathcal{L}_i(\boldsymbol{\theta}, \boldsymbol{\delta}) \Big] \mathrm{d}G(\mathbf{v}_i; \boldsymbol{\delta}) \\
&= \sum_{i=1}^{N} \ln \Big\{ \int_{\mathbb{R}^7} f(in_i|\mathbf{z}_i, \mathbf{v}_i; \boldsymbol{\theta}_{in}) \cdot f(tr_i|\mathbf{z}_i, \mathbf{v}_i, in_i; \boldsymbol{\theta}_{tr}) \\
&\cdot \prod_{t=1}^{T} \big[f(ev_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}; \boldsymbol{\theta}_{ev})^{1-g_{it}} f(di_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}; \boldsymbol{\theta}_{di})^{g_{it}} \\
&\cdot f(out_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it}; \boldsymbol{\theta}_{out})^{s_{it}} \\
&\cdot f(re_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it} = B; \boldsymbol{\theta}_{re})^{1-g_{it}} \\
&\cdot f(dow_{it}|\mathbf{z}_i, \mathbf{v}_i, \Im_{it-1}, ev_{it}, re_{it}; \boldsymbol{\theta}_{dow})^{c_{it}} \big] \Big\} \mathrm{d}G(\mathbf{v}_i; \boldsymbol{\delta}), \qquad (4)
\end{aligned}
$$

where $\mathcal{L}_i(\boldsymbol{\theta}, \boldsymbol{\delta})$ is the individual contribution to the likelihood and $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ are parameters fully characterising the probability density functions with respect to which the sample log-likelihood will be maximised.

Assumption 3 *(Logit and ordered logit probability density functions):*
The probability density functions of both dichotomous and ordered response outcome variables are assumed to have a logit form.

In Subsection 4.2, we clarify in more detail how the explanatory variables and past realisations enter the specification of the logit and ordered logit models of the probability density functions in the log-likelihood function (4). We also explain how we deal with the unobserved heterogeneity distribution and how we allow the unobserved determinants to interact with retention episodes.

4.2 The Empirical Specification

The Initial Conditions

Students start secondary school at different ages, meaning that they have different numbers of years of delay. This is due to a delayed beginning of primary school and/or retention in primary school. In our econometric model, years of delay at the beginning of each secondary school year can affect schooling choices and performances. This variable evolves over time according to the realisation of episodes of retention, which is also one of the outcome variables. As such, years of delay at the beginning of secondary school cannot be assumed to be a nonstochastic starting position for each student. It is very likely to be endogenous since correlated to the unobserved determinants of schooling choices and performances. This results in an initial conditions problem that we solve by specifying an ordered logit model for the years of delay, where unobserved characteristics are allowed to be correlated to those determining future outcomes and choices.

We specify the probability density function of the number of years of delay at the beginning of secondary school as an ordered logit model. This outcome variable takes on the values -1, 0 and 1. We define as $\alpha_{1,in} < \alpha_{2,in}$ the unknown cut points (threshold parameters) and as Λ the logit function. The unobserved heterogeneity component $v_{i,in}$ enters the specification as a shift in the threshold parameters. The probability density function of the initial conditions is:

$$\begin{aligned}
\Pr(in_i = -1|\mathbf{z}_i, v_{i,in}) &= \Lambda(\alpha_{1,in} + v_{i,in} - \mathbf{z}_i'\boldsymbol{\beta}_{in}), \\
\Pr(in_i = 0|\mathbf{z}_i, v_{i,in}) &= \Lambda(\alpha_{2,in} + v_{i,in} - \mathbf{z}_i'\boldsymbol{\beta}_{in}) - \Lambda(\alpha_{1,in} + v_{i,in} - \mathbf{z}_i'\boldsymbol{\beta}_{in}), \\
\Pr(in_i = 1|\mathbf{z}_i, v_{i,in}) &= 1 - \Lambda(\alpha_{2,in} + v_{i,in} - \mathbf{z}_i'\boldsymbol{\beta}_{in}). \qquad (5)
\end{aligned}$$

Students with a higher level of $v_{i,in}$ are less likely to end up into the top category, i.e. to start secondary school one year late.

The Track Choice at the Beginning of Secondary School

The track choice takes value on $\{\mathrm{BSO}, \mathrm{TSO}-, \mathrm{TSO}+, \mathrm{ASO}-, \mathrm{ASO}+\}$. The probability density function of the choice of the hierarchically ordered tracks is:

$$\begin{aligned}
\Pr(tr_i = \mathrm{BSO}|\mathbf{x}_i, in_i, v_{i,tr}) &= \Lambda(\alpha_{1,tr} + v_{i,tr} - \mathbf{x}_i'\boldsymbol{\beta}_{tr} - in_i\gamma_{tr}),\\
\Pr(tr_i = \mathrm{TSO}-|\mathbf{x}_i, in_i, v_{i,tr}) &= \Lambda(\alpha_{2,tr} + v_{i,tr} - \mathbf{x}_i'\boldsymbol{\beta}_{tr} - in_i\gamma_{tr}) - \Lambda(\alpha_{1,tr} + v_{i,tr} - \mathbf{x}_i'\boldsymbol{\beta}_{tr} - in_i\gamma_{tr}),\\
\Pr(tr_i = \mathrm{TSO}+|\mathbf{x}_i, in_i, v_{i,tr}) &= \Lambda(\alpha_{3,tr} + v_{i,tr} - \mathbf{x}_i'\boldsymbol{\beta}_{tr} - in_i\gamma_{tr}) - \Lambda(\alpha_{2,tr} + v_{i,tr} - \mathbf{x}_i'\boldsymbol{\beta}_{tr} - in_i\gamma_{tr}),\\
\Pr(tr_i = \mathrm{ASO}-|\mathbf{x}_i, in_i, v_{i,tr}) &= \Lambda(\alpha_{4,tr} + v_{i,tr} - \mathbf{x}_i'\boldsymbol{\beta}_{tr} - in_i\gamma_{tr}) - \Lambda(\alpha_{3,tr} + v_{i,tr} - \mathbf{x}_i'\boldsymbol{\beta}_{tr} - in_i\gamma_{tr}),\\
\Pr(tr_i = \mathrm{ASO}+|\mathbf{x}_i, in_i, v_{i,tr}) &= 1 - \Lambda(\alpha_{4,tr} + v_{i,tr} - \mathbf{x}_i'\boldsymbol{\beta}_{tr} - in_i\gamma_{tr}),
\end{aligned} \tag{6}$$

where $\mathbf{x}_i \subset \mathbf{z}_i$ due to an exclusion restriction. As exclusion restriction, we use the years of delay at the beginning of primary school. We assume therefore that, at the beginning of secondary school, choices and performances are not affected by the years of delay at the beginning of primary school but just by the years of delay at the start of secondary school, conditional on the other covariates and unobserved heterogeneity.

The Evaluation

At the end of each academic year, pupils receive an evaluation: A, B, or C. As mentioned before, an A allows students to move to the next grade. Students getting a C must resit the grade. Students with a B can decide to downgrade the track if they wish to avoid resitting the grade. The probability density function of the evaluation variable is specified as follows:

$$\begin{aligned}
\Pr(ev_{it} = C|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, v_{i,ev}) &= \Lambda\left[\alpha_{1,ev} + v_{i,ev} - \mathbf{x}_i'\boldsymbol{\beta}_{ev} - \phi_{ev}(in_i, tr_i, \Im_{it-1})\right],\\
\Pr(ev_{it} = B|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, v_{i,ev}) &= \Lambda\left[\alpha_{2,ev} + v_{i,ev} - \mathbf{x}_i'\boldsymbol{\beta}_{ev} - \phi_{ev}(in_i, tr_i, \Im_{it-1})\right]\\
&\quad - \Lambda\left[\alpha_{1,ev} + v_{i,ev} - \mathbf{x}_i'\boldsymbol{\beta}_{ev} - \phi_{ev}(in_i, tr_i, \Im_{it-1})\right],\\
\Pr(ev_{it} = A|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, v_{i,ev}) &= 1 - \Lambda\left[\alpha_{2,ev} + v_{i,ev} - \mathbf{x}_i'\boldsymbol{\beta}_{ev} - \phi_{ev}(in_i, tr_i, \Im_{it-1})\right],
\end{aligned} \tag{7}$$

where $\phi_{ev}(in_i, tr_i, \Im_{it-1})$ is the impact of past outcome variables on future evaluations. We impose some parametric restrictions on the way in which the past is allowed to affect the future. We keep in mind that, from the policy perspective, it is of interest to understand whether and how students' performance is affected by past retention episodes and by past track downgrades. The impact of past outcome variables on future evaluations is modelled as follows:

$$\begin{aligned}\phi_{ev}(in_i, tr_i, \Im_{it-1}) &= \eta_{ev} tr_{it} + \pi_{ev} dow_{it-1} \\ &+ \kappa_{ev} re_{it-1} + \tau_{ev} re_{it-1} \cdot ev_{it-1} + \psi_{ev} tre_{it-1}, \end{aligned} \tag{8}$$

where tr_{it} is the track at the beginning of the t-th academic year, re_{it-1} is an indicator variable equal to 1 if the individual was retained at the end of the previous year (resitting therefore in the current year) and $tre_{it-1} = in_i + \sum_{s=1}^{t-1} re_{is}$ is the total years of delay at the beginning of the t-th academic year. The coefficients κ_{ev} and τ_{ev} capture the transitory effect of retention on the subsequent academic performance, while and ψ_{ev} is the permanent effect. η_{ev} captures track heterogeneity in the ability of the students to get good evaluations. Finally, π_{ev} is the effect of downgrading at the end of the last year on the current schooling achievement.

The School Drop-Out

In Belgium, compulsory education ends on 30 June of the year in which the youth reach the age of 18. From that date onwards, the student is at risk of school drop-out without diploma. Ignoring school drop-out might lead to sample selection attrition as it is not likely to be a random process. We model therefore also the probability of exiting school without the diploma at the end of each year, where the unobserved components determining the school drop-out are allowed to be correlated to the unobserved determinants of the other endogenous processes. In the sequentiality of the events, school drop-out takes place at the end of the academic year, after receiving the evaluation.[8] The school drop-out variable is binary and equal to 1 in case of drop-out. The logit model for pupils at risk of exit is:

$$\begin{aligned}\Pr(out_{it}=1|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, ev_{it}, v_{i,out}) &= \Lambda\big[\alpha_{out} + v_{i,out} + \mathbf{x}_i'\boldsymbol{\beta}_{out} \\ &+ \phi_{out}(in_i, tr_i, \Im_{it-1}, ev_{it})\big], \end{aligned} \tag{9}$$

Similar to Equation (8), the impact of past outcomes on the drop-out probability is

$$\begin{aligned}\phi_{out}(in_i, tr_i, \Im_{it-1}, ev_{it}) &= \eta_{out} tr_{it} + \pi_{out} dow_{it-1} + \omega_{out} ev_{it} \\ &+ \kappa_{out} re_{it-1} + \tau_{out} re_{it-1} \cdot ev_{it-1} + \psi_{out} tre_{it-1}. \end{aligned} \tag{10}$$

[8] Very few students (71, 1.7% of the sample) drop out of school before the end of the academic year. In order to simplify the model and the timing of events, in these cases we advance the drop-out date at the end of the previous academic year, disregarding information on retention and track downgrade of the uncompleted academic year.

Compared to Equation (8), ϕ_{out} has the extra argument, ev_{it}, i.e. the evaluation of the just ended academic year. Under the sequentiality assumption (Assumption 1), ev_{it} is predetermined with respect to the realisation of the drop-out variable. Thereby, it acts as a valid exclusion restriction in the drop-out equation.

The Resitting Choice for B Students

Students getting a B can choose either to resit or to downgrade the track. The choice is binary and, conditional on getting a B, the probability of resitting the grade is specified as a logit model:

$$\Pr(re_{it}=1|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, ev_{it}=B, v_{i,re})=\Lambda\left[\alpha_{re}+v_{i,re}+\mathbf{x}_i'\boldsymbol{\beta}_{re}+\phi_{re}(in_i, tr_i, \Im_{it-1})\right]. \quad (11)$$

The function $\phi_{re}(in_i, tr_i, \Im_{it-1})$ is parametrised as Equation (8):

$$\begin{aligned} \phi_{re}(in_i, tr_i, \Im_{it-1}) &= \eta_{re} tr_{it} + \pi_{re} dow_{it-1} \\ &+ \kappa_{re} re_{it-1} + \tau_{re} re_{it-1} \cdot ev_{it-1} + \psi_{re} tre_{it-1}. \end{aligned} \quad (12)$$

The Track Downgrade

In Belgium, at the beginning of secondary school, students can choose among different tracks characterised by different curricula. This tracking system is aimed at grouping students with similar abilities and preferences. Choosing the right track is important as it will determine future work and education opportunities. However, the initial track choice is not always binding. Students are indeed allowed to switch track at the beginning of a new academic year, although under a set of constraints. The tracks are hierarchically ordered and students can only move from the more general (and more prestigious) tracks to the more specialised and vocationally oriented (and less prestigious) ones. The Belgian system of tracking is therefore often referred to as a 'cascade' system.

We model track transitions by defining a categorical ordered dependent variable for track downgrade. The ordered categories are no downgrade, one-step downgrade and two-step downgrade. They are coded as 0, 1 and 2, respectively.[9] Students in the BSO track are already at the bottom of the cascade and cannot downgrade further. Hence, we model track downgrade only for ASO/TSO students. For BSO students, track downgrade will not give any contribution to the likelihood function. The probability density function

[9] In our dataset, we observe only 48 track transitions of three or more steps. Hence, given the knowledge of a starting point, information on track changes compressed in no downgrade, 1-step downgrade and 2-step downgrade is able to describe almost all the possible track transitions.

of track downgrade for ASO/TSO students is:

$$
\begin{aligned}
\Pr(dow_{it} &= 0|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, ev_{it}, re_{it}, v_{i,dow}) = \\
&\Lambda\big[\alpha_{1,dow} + v_{i,dow} - \mathbf{x}_i'\boldsymbol{\beta}_{dow} - \phi_{dow}(in_i, tr_i, \Im_{it-1}, ev_{it}, re_{it})\big], \\
\Pr(dow_{it} &= 1|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, ev_{it}, re_{it}, v_{i,dow}) = \\
&\Lambda\big[\alpha_{2,dow} + v_{i,dow} - \mathbf{x}_i'\boldsymbol{\beta}_{dow} - \phi_{dow}(in_i, tr_i, \Im_{it-1}, ev_{it}, re_{it})\big] \\
&- \Lambda\big[\alpha_{1,dow} + v_{i,dow} - \mathbf{x}_i'\boldsymbol{\beta}_{dow} - \phi_{dow}(in_i, tr_i, \Im_{it-1}, ev_{it}, re_{it})\big], \\
\Pr(dow_{it} &= 2|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, ev_{it}, re_{it}, v_{i,dow}) = \\
&1 - \Lambda\big[\alpha_{2,dow} + v_{i,dow} - \mathbf{x}_i'\boldsymbol{\beta}_{dow} - \phi_{dow}(in_i, tr_i, \Im_{it-1}, ev_{it}, re_{it})\big]. \quad (13)
\end{aligned}
$$

The function $\phi_{dow}(in_i, tr_i, \Im_{it-1}, ev_{it}, re_{it})$ is linearly specified as follows:

$$
\begin{aligned}
\phi_{dow}(in_i, tr_i, \Im_{it-1}) &= \eta_{dow} tr_{it} + \pi_{dow} dow_{it-1} + \omega_{dow} ev_{it} + \xi_{dow}(1 - re_{it}) \\
&+ \kappa_{dow} re_{it-1} + \tau_{dow} re_{it-1} \cdot ev_{it-1} + \psi_{dow} tre_{it-1}, \quad (14)
\end{aligned}
$$

where ξ_{dow} is the effect of not being retained at the end of the academic year on the probability of downgrading the track.

The Diploma Equation

In the last grade of the track (6th grade for ASO/TSO and 7th grade for BSO), students do not receive an evaluation with marks A, B, or C. If they succeed, they simply get the diploma. If they fail, they have to resit the last grade.[10] The performance variable of the last grade of secondary school is therefore binary. We specify the probability of success, which implies getting the secondary school diploma, as a logit model:

$$
\Pr(di_{it} = 1|\mathbf{x}_i, in_i, tr_i, \Im_{it-1}, v_{i,di}) = \Lambda\big[\alpha_{di} + v_{i,di} + \mathbf{x}_i'\boldsymbol{\beta}_{di} + \phi_{di}(in_i, tr_i, \Im_{it-1})\big]. \quad (15)
$$

The function $\phi_{di}(in_i, tr_i, \Im_{it-1})$ has a linear form as in Equation (8):

$$
\begin{aligned}
\phi_{di}(in_i, tr_i, \Im_{it-1}) &= \eta_{di} tr_{it} + \pi_{di} dow_{it-1} \\
&+ \kappa_{di} re_{it-1} + \tau_{di} re_{it-1} \cdot ev_{it-1} + \psi_{di} tre_{it-1}. \quad (16)
\end{aligned}
$$

[10]Students failing the last grade of ASO/TSO can also choose to switch to grade 6 of the BSO track, which is taken into account in the model.

The Unobserved Heterogeneity Distribution

In order to maximise the log-likelihood function in (4), we need to assign some parametric form to the joint distribution of the unobserved heterogeneity component $\mathbf{v}_i \equiv (v_{i,in}, v_{i,tr}, v_{i,ev}, v_{i,out}, v_{i,re}, v_{i,dow}, v_{i,di})$. In order to avoid too strict parametric assumptions, we follow Heckman and Singer (1984) and assume that $G(\mathbf{v}_i)$ is discrete with a finite and, *a priori*, unknown number M points of support. However, estimating our model with a seven-dimensional discrete distribution would be computational demanding. Our outcome variables belong to three types: i) the initial conditions; ii) schooling achievements (evaluation and diploma acquisition); iii) educational choices (track choice, downgrade choice, resitting decision in case of B and drop-out decision). In order to reduce the estimation complexity of the model, we reduce the dimension of $\mathbf{v}_i$ to three by one-factor loading specifications: $v_{i,di} = \delta_{di} \cdot v_{i,ev}$, $v_{i,tr} = \delta_{tr} \cdot v_{i,dow}$, $v_{i,out} = \delta_{out} \cdot v_{i,dow}$ and $v_{i,re} = \delta_{re} \cdot v_{i,dow}$.

On the basis of Monte Carlo simulations for treatment effects in duration models, Gaure et al. (2007) find that the number of the points of support is best chosen by minimising the Akaike Information Criterion (AIC). We follow this recommendation. The probabilities associated to the points of support sum to one and are, $\forall\ m = 1, \ldots, M$, denoted by

$$p^m = \Pr(v_{i,in} = v^m_{i,in}, v_{i,ev} = v^m_{i,ev}, v_{i,dow} = v^m_{i,dow}) \equiv \Pr(\mathbf{v}_i = \mathbf{v}^m_i)$$

and specified as logistic transforms:

$$p^m = \frac{\exp(\lambda^m)}{\sum_{g=1}^{M} \exp(\lambda^g)} \quad \text{with} \quad m = 1, \ldots, M \quad \text{and} \quad \lambda_M = 0.$$

The sample log-likelihood function in Equation (4) can be rewritten as

$$\ell(\boldsymbol{\theta}, \boldsymbol{\delta}) = \sum_{i=1}^{N} \ln \Bigg[\sum_{m=1}^{M} p^m \mathcal{L}_{im}(\boldsymbol{\theta}, \boldsymbol{\delta}) \Bigg], \tag{17}$$

where $\mathcal{L}_{im}(\boldsymbol{\theta}, \boldsymbol{\delta})$ is the individual contribution to the likelihood function if the individual is of type m.

During the empirical analysis, we also use an alternative specification of the unobserved heterogeneity support points. In this alternative specification, we allow the points of support of v^m_{ev} and v^m_{dow} to interact with lagged retention and cumulated retention for each $m = 1, \cdots, M$:

$$v^m_{it,ev} = v^m_{i,ev}(1 + \psi_{ev} re_{it-1} + \zeta_{ev} tre_{it-1}) \quad (18)$$
$$v^m_{it,dow} = v^m_{i,dow}(1 + \psi_{dow} re_{it-1} + \zeta_{dow} tre_{it-1}). \quad (19)$$

By doing so, we allow the transitory and permanent effect of grade retention to be heterogeneous across unobserved determinants of preferences and choices. In other words, the points of support become time-varying, depending on the retention realisation. These time-varying components have to be plugged into models (7), (9), (11), (13) and (15).

4.3 Identification

The identification of the interrelated dynamics between grade retention, track mobility, and schooling attainment is obtained by addressing some key challenges. In this subsection we summarise the aforementioned characteristics of our model that induce this identification.

First, educational achievements and choices are likely to be determined by a set of unobserved determinants, e.g. behavioural and cognitive skills, with unknown correlation structure. In order to disentangle the pure effects of past educational outcomes on future ones from the spurious effects determined by unobserved abilities, we take into account the presence of unobserved heterogeneity by semi-parametric maximum likelihood techniques (Heckman and Singer, 1984; Mroz, 1999). The identification of the unobserved heterogeneity distribution is based on multiple observations per student of the same processes.

Second, the imposed sequencing of schooling achievements and choices makes some of the outcome variables determinants of later outcomes within each academic year. This generates predetermined exclusion restrictions which are used to identify the interrelated dynamics of schooling achievements and choices.

Third, at the start of secondary school pupils have already different years of delay due to retention episodes either in kindergarten or in primary school. If we assume that grade retention affects future outcome variables, we have an initial conditions problem. We solve for initial conditions by adding an equation for the years of delay at the beginning of secondary school, which depends on unobserved heterogeneity and an exclusion restriction (Heckman, 1981). As exclusion restriction, we use the number of years of delay at the start of primary school. We assume therefore that once we control for the years of delay at the start of secondary school, the years of delay at the start of primary school do not affect secondary school performances and choices.

Fourth, as pointed out by Fruehwirth et al. (2011), the effect of grade retention might be heterogeneous and vary by students' unobserved abilities. We allow therefore the effect of past retention episodes to vary across different levels of the unobserved determinants by imposing a specific functional forms on the interaction effect.

Finally, there might be sample selection attrition induced by students dropping out of secondary school. We model therefore also the probability of exiting school without the diploma at the end of each year, where the unobserved components determining the school drop-out are allowed to be correlated to the unobserved determinants of the other endogenous processes. The loading factor structure of the unobserved heterogeneity component and the fact that some students are at risk of drop-out for more than one year are of help in identifying the attrition equation.

4.4 Partial Observability of Tracks at the Start of Secondary School

As mentioned in Section 3, at the beginning of secondary school, we have only partial information about the school track choice. We only know whether students are in the vocational track (BSO) or not (ASO/TSO). Only starting from grade 2 we have detailed information on courses of study and we can group students into the five tracks. However, the cascade system of the institutional set-up jointly with the track position and track mobility of each student in subsequent grades convey some information about the possible starting track. For example, students who are in ASO+ in grade 2, surely were also in ASO+ in grade 1, as track upgrading is not allowed. For the same reason, students in ASO− in grade 2 were not in TSO and BSO tracks in grade 1. We modify the likelihood function to take into account the partial observability of the track at the beginning of secondary school: we integrate over the possible tracks in grade 1, given future information about tracks and mobility. This is similar to the strategy used by Mroz and Picone (2011) to solve the partial observability of the time in which persons with diabetes progress to the next disease stage.

To show in what direction we modify the likelihood function and keep the notation simple, we rewrite the density in Equation (1) by ignoring the conditioning on the observed and unobserved covariates and the individual subscript i, yielding

$$f(in, tr, \mathbf{Y}) = f(in)f(tr|in)f(\mathbf{Y}|tr, in). \tag{20}$$

We assume that the probability of being in each track at the beginning of secondary school is related to the information we have in the future about tracks, mobility choices and performances. Denote by $f(tr|in, \mathbf{Y})$ this probability density function. If we integrate

Equation (20) over the possible tracks, we get

$$f(in, \mathbf{Y}) = f(in) \int f(tr|in) f(\mathbf{Y}|tr, in) f(tr|in, \mathbf{Y}) \mathrm{d}tr. \tag{21}$$

Once we parametrise $f(tr|in)$ and $f(\mathbf{Y}|tr, in)$, like we did in Subsection 4.2, we imply a particular parametrisation of $f(tr|in, \mathbf{Y})$:

$$\begin{aligned} f(tr|in, \mathbf{Y}) &= \frac{f(\mathbf{Y}|tr, in) f(tr|in)}{f(\mathbf{Y}|in)} \\ &= \frac{f(\mathbf{Y}|tr, in) f(tr|in)}{\int f(tr, \mathbf{Y}|in) \mathrm{d}tr} \\ &= \frac{f(\mathbf{Y}|tr, in) f(tr|in)}{\int f(\mathbf{Y}|tr, in) f(tr|in) \mathrm{d}tr}. \end{aligned} \tag{22}$$

Both the numerator and the denominator of Equation (22) depend indeed on the probability density functions that we have already parametrised in Subsection 4.2. Substituting Equation (22) into Equation (21) yields

$$f(in, \mathbf{Y}) = f(in) \int \frac{f(tr|in)^2 f(\mathbf{Y}|tr, in)^2}{\int f(\mathbf{Y}|s, in) f(s|in) \mathrm{d}s} \mathrm{d}tr. \tag{23}$$

Since tracks take value on 5 categories, the integrals in Equation (23) are just sums over the 5 possible realisations. The individual contribution to the likelihood function in Equation (3) and the sample log-likelihood function in Equation (4) are modified along the lines dictated by Equation (23).

5 Estimation Results

The econometric model is made up of seven equations. The estimation results of the coefficients of each equation are reported and commented in the next subsections. We display estimation results of three different model specifications: without unobserved heterogeneity, with time-invariant unobserved heterogeneity and with time-invariant unobserved heterogeneity interacted with lagged retention and the cumulated years of delay. Table 3 reports the estimation results of the probability masses of the discrete unobserved heterogeneity distribution. The number of points of support are chosen by minimising the AIC. For both specifications controlling for unobserved heterogeneity the resulting number is 6. The preferred model according to the AIC is the one that encompasses the interactions between the unobserved heterogeneity and lagged retention and cumulated retention.

Table 3: Estimated Probability Masses of the Discrete Unobserved Heterogeneity Distribution and Other Statistics

	Without unobserved heterogeneity (1)		With time-invariant unobserved heterogeneity (2)			With time-invariant and time-variant unobserved heterogeneity (3)		
	Coeff.	Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.
Unobserved heterogeneity probability masses (λ_6 is normalized to 0)								
λ_1	–	–	-1.595	***	0.458	-1.545	***	0.467
λ_2	–	–	-2.929	***	0.338	-2.583	***	0.279
λ_3	–	–	-1.171	***	0.304	-1.161	***	0.253
λ_4	–	–	0.256		0.267	0.290		0.281
λ_5	–	–	0.099		0.262	0.163		0.261
Resulting probability masses								
p_1	–		0.051			0.052		
p_2	–		0.013			0.018		
p_3	–		0.078			0.076		
p_4	–		0.326			0.325		
p_5	–		0.279			0.286		
p_6	–		0.252			0.243		
Log-likelihood	-22,380.5		-22,222.2			-22,194.2		
AIC/N	10.679		10.615			10.604		
Number of parameters	120		144			148		
Number of pupils (N)	4,214		4,214			4,214		

5.1 Initial Conditions: Years of Delay at the Beginning of Secondary School

Table 4 reports the estimation results of the ordered logit model for the years of delay at the beginning of secondary school. The estimation results of the initial conditions equation are very stable across the three model specifications. We find that the years of delay at the start of primary school, i.e. the exclusion restriction, strongly and positively affects the probability of starting secondary school with delay. The relative age determined by birth date has a significant negative effect on the years of delay at the beginning of secondary school: the later in the year the kid was born, the higher the probability that (s)he will cumulate years of delay. This evidence is consistent with those in Bedard and Dhuey (2006), Fredriksson and Öckert (2006), Hámori (2007), McEwan and Shapiro (2008), Strøm (2004) and Altwicker-Hámori and Köllő (2012), who find that school starting age has a positive effect on several measures of academic performance.

Parents' education has a significant impact on years of delay at the start of secondary school, especially mother's education: the higher the education of the mother of the student, the lower the probability that the pupil begins secondary education with delay. The number of siblings is also a significant determinant of the years of delay: in line with the

Table 4: Estimation Results of the Initial Conditions Equation: Years of Delay at the Beginning of Secondary School

	Without unobserved heterogeneity (1)			With time-invariant unobserved heterogeneity (2)			With time-invariant and time-variant unobserved heterogeneity (3)		
Variables	Coeff.		Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.
Years of delay at the start of primary school	5.317	***	0.253	10.275	***	0.593	10.505	***	0.621
Female	-0.137		0.164	-0.221		0.224	-0.225		0.224
Cohort 1980	-0.217		0.171	-0.215		0.228	-0.198		0.228
Calendar day of birth/100	1.549	***	0.141	1.960	***	0.242	1.965	***	0.243
Father's education/10	-0.634	**	0.302	-0.774	*	0.421	-0.751	*	0.422
Mother's education/10	-1.484	***	0.312	-1.636	***	0.465	-1.696	***	0.460
Number of siblings – Reference: No siblings									
1 sibling	0.232		0.254	0.340		0.344	0.356		0.342
2 siblings	0.072		0.277	0.177		0.374	0.204		0.375
3 or more	0.744	***	0.285	1.029	**	0.401	1.046	***	0.400
Unobserved heterogeneity support points									
v_2	–		–	-0.871		8.336	-1.587		7.551
v_3	–		–	-10.582	***	1.875	-11.184	***	1.732
v_4	–		–	-1.313		1.530	-1.371		1.442
v_5	–		–	-10.419	***	1.952	-10.981	***	1.772
v_6	–		–	-1.997		2.362	-2.753		2.103

effect on test scores in Hámori (2007), we find that pupils with more than two siblings have a higher probability of starting secondary school with delay. This effect might be explained by the fact that in larger families parents have less time to dedicate to each child. It might also be that the number of siblings capture particular social and cultural family background.

5.2 Track Choice at the Beginning of Secondary School

Table 5 displays the estimation results of the equation for the track choice at the beginning of secondary school. The tracks are hierarchically ordered from the bottom (BSO) to the top (ASO+). The years of delay at the beginning of secondary school significantly reduce the probability of choosing the ASO+ track and increase the probability of preferring the vocational track. When we control for unobserved heterogeneity the impact of past schooling performances captured by the years of delay at the start of secondary school gets smaller in absolute value. This means that part of the effect is spurious: unobserved characteristics, like ability and intelligence, jointly determine the probability of starting late secondary school and the track choice. Once we net out the spurious negative correlation between unobserved ability and the probability of starting late secondary school, the coefficient of the impact is reduced in size.

All the other regressors are highly significant in explaining the school track choice.

Table 5: Estimation Results of the Track Choice Equation

	Without unobserved heterogeneity (1)			With time-invariant unobserved heterogeneity (2)			With time-invariant and time-variant unobserved heterogeneity (3)		
	Coeff.		Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.
Years of delay at the start of secondary school	-1.277	***	0.136	-0.804	***	0.174	-0.785	***	0.176
Female	0.318	***	0.058	0.448	***	0.077	0.428	***	0.076
Cohort 1980	-0.150	***	0.058	-0.238	***	0.076	-0.250	***	0.075
Calendar day of birth/100	-0.166	***	0.029	-0.273	***	0.040	-0.275	***	0.039
Father's education/10	1.651	***	0.109	2.262	***	0.155	2.233	***	0.154
Mother's education/10	1.741	***	0.118	2.310	***	0.166	2.283	***	0.165
Number of siblings – Reference: No siblings									
1 sibling	-0.111		0.090	-0.174		0.118	-0.178		0.117
2 siblings	-0.243	**	0.099	-0.351	***	0.130	-0.354	***	0.128
3 or more	-0.425	***	0.110	-0.584	***	0.142	-0.585	***	0.140
Unobserved heterogeneity loading factor									
Loading factor	–		–	-6.347	***	2.063	-6.403	***	2.144

Girls and pupils from highly educated parents are less likely to choose the vocational track and more likely to get into ASO+. The gender effect might be induced by gender heterogeneous preferences for vocational/technical tracks but also influenced by the socio-cultural environment, the performance expectations and their interaction.[11]

Both mother's and father's education strongly push up the probability of choosing the highest track (ASO+) and discourages the vocational track (BSO), meaning that parents take influence on the education of their children. This is a quite common association found in the educational research literature. See Haveman and Wolfe (1995) for a review of the literature on intergenerational mobility with respect to education and, among others, Bratti et al. (2012), Dustmann (2004) and Falter et al. (2011) for more recent findings on the effect of parental background on pupils' track choices. Also the family structure has an impact on track choice: the larger the number of siblings the higher the probability of choosing the vocational track (BSO). Finally, the younger the pupil, the higher the probability of choosing a lower track.

5.3 Evaluation at the End of the Academic Year

There are several studies in the educational research literature aimed at understanding whether grade retention has a positive or a negative impact on subsequent academic performances. See for instance the literature review in Xia and Kirby (2009) and the meta-analysis in Jimerson (2001). The conclusions are not uncontroversial. Most of the studies find a negative relationship between retention and subsequent academic achievement.

[11] Guiso et al. (2008) show that the more the culture is gender-equal, the better the girls score in math.

However, if the analyst cannot control for all the determinants of grade retention and subsequent performances the estimate will be biased due to a selection bias. Innate ability, intelligence, cognitive skills and commitment to work are determinants of both grade retention and future educational achievements. If they are not properly taken into account, the impact of grade retention will be spurious and biased downwards.

In existing studies, the identification of the causal effect of grade retention mostly relies on controlling for confounding factors or on matching students on the basis of a set of observable characteristics. A few studies address the selection bias by instrumental variables (IV) relying on shifts and discontinuities determined by retention policies (Fruehwirth et al., 2011; Eide and Showalter, 2001; Greene and Winters, 2007; Jacob and Lefgren, 2004, 2009; Manacorda, 2012) or on the independence between the instrument and the selection variable, conditional on the outcome (D'Haultfœuille, 2010). For the French speaking region of Belgium, Belot and Vandenberghe (2011) exploit a reform which reintroduced the possibility of retention in the first grade of secondary education, finding no impact on academic performance.

In this study, we do not need to assume that we are controlling for all factors determining both the treatments (retention and track mobility) and the outcomes (some measures of subsequent performance), like in the matching literature. Moreover, we do not need a valid IV or an exclusion restriction. We rather exploit the longitudinal dimension of the dataset and the availability of multiple observations per student of the achievement and choice variables. This rich information allows us to flexibly identify the unobserved heterogeneity distribution and the correlation between the unobserved determinants of the performance outcomes (e.g. evaluation), of the choices (e.g. retention and track downgrade) and of the initial conditions (the years of delay at the beginning of secondary school).

The estimation results of the evaluation ordered logit equation are displayed in Table 6. In all three specifications, the transitory impact of retention (the coefficient of lagged retention) has a positive impact on the next evaluation. Hence, *ceteris paribus*, pupils who are resitting the grade are less likely to get a C and thereby to be retained again, than students who are not resitting. In contrast, based on model (2) controlling for time-invariant unobserved heterogeneity, for the permanent effect (the coefficient of total years of delay) we get a negative effect. In model (3), which allows the retention effects to be heterogeneous across different levels of the unobserved component v_{ev}, however, also the permanent effect is positive. This means that an episode of grade retention will also have a positive effect on the evaluation of all the next academic years. The evidence of a positive impact of grade retention on future schooling achievements contrasts with prior research. Two exceptions are D'Haultfœuille (2010) and Jacob and Lefgren (2004), who found that in the US and France, respectively, grade retention has a positive short-term effect on schooling performance. The former research based identification on a new method for models with endogenous selection and exploited the independence between an instrument and the selection variable, conditional on the outcome. The latter exploited a discontinuity generated by a school reform.

Two points are worthy of mention about the estimation results when moving to model (3). First, the permanent effect of grade retention (the coefficient of total years of delay) switches sign, from negative to positive. Hence, when we do not take into account that pupils might react differently to retention by abilities, like cognitive skills, intelligence and commitment, the permanent effect of retention is underestimated. Second, the interaction between the unobserved heterogeneity component and total years of delay is significantly negative: if the unobserved component is small enough ($v_{ev} << 0$), for instance if the student is very smart, grade retention generates a net negative permanent effect. Less able pupils ($v_{ev} > 0$) will instead be permanently, as well as momentarily, favoured by grade retention. The fact that more able students might be permanently penalised by an episode of grade retention suggests that psychological costs dominate possible benefits. For less able pupils the psychological costs might instead be dominated by the benefit of having more time to develop the knowledge and emotional maturity required at each educational grade. Also Fruehwirth et al. (2011) find that the retention effect varies by the abilities of pupils retained in kindergarten in the US. However, in contrast to our results, they find that lower able pupils are more negatively affected by grade retention. We conclude that when assessing the effectiveness of grade retention, one should carefully consider the heterogeneity in responses to grade retention by unobservable behavioural and cognitive abilities.

In the top track (ASO+) and in the vocational track (BSO), it is easier to get top evaluations. Pupils downgrading track are more likely to get good evaluations in the next academic year. Although there might be negative effects induced by changing peers and sometimes school, students who come from a higher track are likely to have an excess of knowledge relatively to the new track-level requirements, so they succeed more easily.

About the impact of exogenous regressors, the results are as expected. Girls perform better than boys, as it is generally found in the educational literature.[12] Parents' education is positively associated to the probability of getting an A. Pupils in larger families are more likely to perform worse.

[12] See for instance the results in Van Houtte (2004) for Flanders.

Table 6: Estimation Results of the Evaluation Equation

	Without unobserved heterogeneity (1)			With time-invariant unobserved heterogeneity (2)			With time-invariant and time-variant unobserved heterogeneity (3)		
	Coeff.		Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.
Exogenous variables									
Female	0.423	***	0.045	0.504	***	0.055	0.467	***	0.054
Cohort 1980	-0.055		0.045	-0.083		0.053	-0.103	**	0.052
Calendar day of birth/100	0.026		0.022	-0.013		0.026	-0.012		0.026
Father's education/10	0.215	***	0.083	0.473	***	0.103	0.432	***	0.102
Mother's education/10	0.200	***	0.088	0.481	***	0.110	0.423	***	0.108
Number of siblings – Reference: No siblings									
1 sibling	-0.015		0.069	-0.036		0.081	-0.035		0.081
2 siblings	-0.143	*	0.074	-0.212	**	0.088	-0.208	**	0.089
3 or more	-0.161	*	0.085	-0.233	**	0.102	-0.231	**	0.101
Grade – Reference: Grade 1									
Grade 2	-1.747	***	0.127	-1.532	***	0.138	-1.529	***	0.137
Grade 3	-1.597	***	0.135	-1.414	***	0.144	-1.413	***	0.144
Grade 4	-1.654	***	0.129	-1.537	***	0.137	-1.523	***	0.138
Grade 5	-1.494	***	0.134	-1.438	***	0.144	-1.414	***	0.144
Grade 6	-1.225	***	0.222	-1.255	***	0.240	-1.186	***	0.238
Time-varying endogenous variables									
Track – Reference: BSO									
ASO+	0.964	***	0.126	0.167		0.186	0.298		0.187
ASO−	-0.541	***	0.080	-1.150	***	0.126	-1.058	***	0.122
TSO+	-0.681	***	0.093	-1.063	***	0.122	-1.028	***	0.120
TSO−	-0.907	***	0.078	-1.187	***	0.101	-1.174	***	0.100
Total years of delay	-0.522	***	0.055	-0.476	***	0.076	0.370	**	0.146
Lagged retention	1.369	***	0.138	1.419	***	0.155	1.033	***	0.367
Lag B if retained last year	0.866	***	0.303	0.782	**	0.311	0.767	**	0.330
Lag A if not retained last year	0.958	***	0.090	0.675	***	0.105	0.662	***	0.106
Downgrade at the end of previous year – Reference: No downgrade									
1-step downgrade	0.597	***	0.098	0.443	***	0.106	0.474	***	0.106
2-step downgrade	0.916	***	0.170	0.648	***	0.179	0.713	***	0.182
Unobserved heterogeneity									
Unobserved heterogeneity support points (v_1 is normalized to 0)									
v_2	–		–	1.876	***	0.273	1.632	***	0.221
v_3	–		–	0.401	***	0.173	0.245	*	0.140
v_4	–		–	-0.111		0.105	-0.088		0.070
v_5	–		–	-0.876	***	0.192	-1.113	***	0.211
v_6	–		–	-1.330	***	0.239	-1.295	***	0.249
Interactions of unobserved heterogeneity support points with retention variables									
Interaction with total years of delay	–		–	–		–	-0.279	***	0.040
Interaction with lagged retention	–		–	–		–	0.114		0.114

5.4 Resitting Decision for B Students

Students getting a B can choose either to resit or to downgrade the track. The estimation results of the resitting equation for B students are reported in Table 7. Few regressors are significant. The higher the education of the father, the higher the probability that the pupil will prefer to resit instead of downgrading the track. The social status of the father of the pupil is therefore not only a determinant of schooling success, but also of resitting/downgrading choices. The lagged retention indicator has a significant negative impact on the probability of choosing retention. This means that retained students who get a B are more likely to downgrade than to resit again compared to non-retained students. The cost of losing an academic year seems therefore to be increasing with the number of times students resit the same grade.

Table 7: Estimation Results of the Resitting Decision for B Students

	Without unobserved heterogeneity (1)			With time-invariant unobserved heterogeneity (2)			With time-invariant and time-variant unobserved heterogeneity (3)		
	Coeff.		Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.
Exogeneous variables									
Female	-0.135		0.164	-0.147		0.168	-0.143		0.167
Cohort 1980	0.024		0.167	0.030		0.169	0.028		0.169
Calendar day of birth/100	0.123		0.077	0.128		0.080	0.127		0.080
Father's education/10	0.716	**	0.298	0.701	**	0.321	0.707	**	0.322
Mother's education/10	0.160		0.319	0.145		0.331	0.153		0.332
1 or more siblings†	0.055		0.259	0.053		0.260	0.054		0.262
Grade – Reference: Grade 2§									
Grade 3	0.573	**	0.227	0.544	**	0.229	0.542	**	0.232
Grade 4	0.722	***	0.196	0.712	***	0.208	0.712	***	0.210
Grade 5	0.969		1.016	0.776		1.030	0.775		1.077
Time-varying endogenous variables									
Track – Reference: BSO									
ASO+	-0.556		0.514	-0.628		0.649	-0.656		0.656
ASO−	0.432		0.322	0.269		0.359	0.269		0.359
TSO+	0.992	***	0.356	0.805	**	0.368	0.810	**	0.369
TSO−	1.144	***	0.326	0.955	***	0.336	0.963	***	0.336
Total years of delay	-0.384		0.270	-0.409		0.281	-0.278		0.322
Lagged retention†	-0.957		0.668	-0.984		0.687	-1.638	**	0.753
Downgrade in the previous year†	0.076		0.323	0.080		0.327	0.070		0.332
Unobserved heterogeneity loading factor									
Loading factor	–		–	-0.281		0.761	-0.193		0.806

† Due to the small number of students getting a B, there is not enough variation to be finer in distinguishing between number of siblings and the number of downgrade steps at the end of the previous academic year. For the same reason, we could not identify the effect of the interactions between lagged retention and lagged evaluation.
§ Nobody resits grade 1 and therefore the reference category is grade 2.

5.5 Track Downgrade

The estimation results of the ordered logit model for track downgrade are reported in Table 8. A positive coefficient means that the corresponding regressor has a positive impact on the probability of making a two-step downgrade and a negative impact on the probability of remaining in the same track.

Parents' education has a negative effect on track downgrade. This evidence, jointly with the finding that the higher the education of the father the higher the probability that students getting a B will prefer to resit instead of downgrading, are in line with the predictions of sociological theories claiming that educational choices are influenced by social status maintenance and structural risk aversion. In a society where education is an investment good for social status upgrade, more advantaged families (higher educated parents) might have a greater incentive to invest in their children's education in order to preserve their advantage (Thurow, 1972). Moreover, higher education might become a social norm which children are persuaded to follow under the pressure of their family and peers (Boudon, 1974). Although also families in less advantaged class positions might invest in their children's education to give them a chance to raise their social and economic position, the failure of getting an education degree for a student from a lower educated family is likely to have more serious consequences than those for a student from families with larger resources (Goldthorpe, 1996; Breen and Goldthorpe, 1997). Since easier and/or shorter educational tracks minimise the risk of failing and entering the labour market without any (vocational) degree (Hartlaub and Schneider, 2012), it is not surprising to find students with lower educated parents to be more likely to choose the vocational track and to prefer downgrading to resitting.

The higher the current track, the higher the probability of downgrading it. Furthermore, the evaluation obtained at the end of the current academic year is a strong determinant of downgrading. As expected, students getting an A are less likely to downgrade than students getting a B and, above all, a C. Students who have experienced a track change are less likely to downgrade in the following year, meaning that downgrading stabilises the track pathways of students. Finally, the total years of delay and ending the year without the need to resit the next one positively affect the probability of track downgrade.

5.6 School Drop-Out without Diploma

Table 9 reports the estimation results of the drop-out equation. First, girls and younger students are less likely to drop out of secondary school without a diploma.[13] Second, the

[13] Eide and Showalter (2001) find the same gender difference in drop-out rates in the US.

Table 8: Estimation Results of the Track Downgrade for ASO/TSO Students

	Without unobserved heterogeneity (1)			With time-invariant unobserved heterogeneity (2)			With time-invariant and time-variant unobserved heterogeneity (3)		
	Coeff.		Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.
Exogenous variables									
Female	0.035		0.056	-0.004		0.058	0.002		0.058
Cohort 1980	-0.041		0.056	-0.026		0.056	-0.024		0.056
Calendar day of birth/100	-0.003		0.027	0.019		0.028	0.019		0.027
Father's education/10	-0.600	***	0.105	-0.757	***	0.116	-0.748	***	0.116
Mother's education/10	-0.416	***	0.114	-0.578	***	0.125	-0.568	***	0.124
Number of siblings – Reference: No siblings									
1 sibling	-0.099		0.082	-0.091		0.084	-0.090		0.083
2 siblings	-0.161	*	0.091	-0.141		0.092	-0.140		0.092
3 or more	-0.092		0.106	-0.063		0.109	-0.066		0.109
Grade – Reference: Grade 1									
Grade 2	2.544	***	0.131	2.454	***	0.133	2.464	***	0.133
Grade 3	1.318	***	0.136	1.237	***	0.137	1.243	***	0.138
Grade 4	1.851	***	0.131	1.814	***	0.132	1.818	***	0.133
Time-varying endogenous variables									
Track – Reference: TSO−									
ASO+	1.193	***	0.104	1.747	***	0.176	1.712	***	0.175
ASO−	0.369	***	0.093	0.641	***	0.108	0.608	***	0.107
TSO+	0.674	***	0.105	0.799	***	0.109	0.778	***	0.109
Current evaluation – Reference: C									
A	-4.297	***	0.197	-4.292	***	0.200	-4.299	***	0.201
B	-0.721	***	0.183	-0.727	***	0.187	-0.732	***	0.187
Total years of delay	0.233	***	0.090	0.209	**	0.092	0.310	**	0.158
Lagged retention	0.414	**	0.198	0.445	**	0.203	-0.121		0.346
No current retention	2.216	***	0.182	2.232	***	0.186	2.232	***	0.186
Lag B if retained last year	-0.541	*	0.295	-0.545	*	0.299	-0.534	*	0.304
Lag A if not retained last year	0.105		0.125	0.119		0.127	0.127		0.127
Downgrade at the end of previous year – Reference: No downgrade									
1-step downgrade	-0.428	**	0.167	-0.290	*	0.170	-0.292	*	0.171
2-step downgrade	-1.117	***	0.367	-0.940	**	0.385	-0.950	**	0.388
Unobserved heterogeneity									
Unobserved heterogeneity support points (v_1 is normalized to 0)									
v_2	–		–	-1.146	***	0.381	-1.073	***	0.368
v_3	–		–	-1.019	***	0.343	-0.990	***	0.344
v_4	–		–	-0.917	***	0.314	-0.889	***	0.316
v_5	–		–	-0.894	***	0.308	-0.880	***	0.313
v_6	–		–	-0.478	***	0.207	-0.464	**	0.210
Interactions of unobserved heterogeneity support points with retention variables									
Interaction with total years of delay	–		–	–		–	0.041		0.049
Interaction with lagged retention	–		–	–		–	-0.237	**	0.102

higher parents' education, the lower the propensity to drop-out. Third, students reaching the final grade, therefore closer to the target, or getting an A are less likely to drop out without the diploma. Last, BSO students have a significantly higher probability of not completing secondary school. This finding might be explained by the fact that BSO students' opportunity cost of not getting the diploma might be lower than the one of ASO/TSO students for at least two reasons. First, students in vocational tracks might access the labour market without the diploma in specialised/blue collar jobs more easily than similar ASO/TSO students because of the specific human capital they acquired in the BSO track. Second, BSO students might be less interested in enrolling in tertiary education.

Table 9: Estimation Results of the Drop-Out Equation

	Without unobserved heterogeneity (1)			With time-invariant unobserved heterogeneity (2)			With time-invariant and time-variant unobserved heterogeneity (3)		
	Coeff.		Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.
Exogenous variables									
Female	-0.183		0.144	-0.489	**	0.200	-0.366	**	0.183
Cohort 1980	0.088		0.138	0.103		0.178	0.156		0.169
Calendar day of birth/100	-0.218	***	0.067	-0.237	***	0.085	-0.218	***	0.082
Father's education/10	-0.454	*	0.270	-1.008	***	0.365	-0.917	***	0.348
Mother's education/10	-0.158		0.281	-0.830	**	0.375	-0.725	**	0.348
1 or more siblings	0.000		0.194	0.075		0.260	0.010		0.241
Final grade†	-3.267	***	0.288	-2.353	***	0.343	-2.748	***	0.353
Time-varying endogenous variables									
BSO†	2.138	***	0.248	2.100	***	0.349	1.992	***	0.300
Current evaluation – Reference: C									
A	-1.450	***	0.228	-0.738	**	0.300	-1.079	***	0.281
B	0.215		0.543	0.575		0.637	0.284		0.645
Total years of delay	-0.036		0.106	-0.214		0.140	-0.211		0.137
Lagged retention†	0.231		0.236	-0.123		0.283	-0.000		0.280
Downgrade in the previous year†	-0.560	*	0.320	-0.555		0.407	-0.541		0.394
Unobserved heterogeneity loading factor									
Loading factor	–		–	-18.430	**	7.604	-16.696	**	6.931

† As students can drop-out of school without the diploma only when they turn 18 years old, the sample at risk of exit is small and there is not enough variation to distinguish between different tracks, grades and the number of downgrade steps at the end of the previous academic year. For the same reason, we could not identify the effect of the interactions between lagged retention and lagged evaluation.

5.7 Secondary School Graduation

Table 10 reports the estimation results of the diploma equation, i.e. the impact of covariates on the probability of getting the diploma once students make it to the last grade of their track.[14] Once again we find that girls and pupils in smaller families perform better and that they are significantly more likely to get the diploma. Parents's education is positively correlated to the probability of getting the diploma, although only the impact of mother's education is significantly different from zero. Finally, once students are in the last grade, the transitory and the permanent retention effects are positive but not significantly different from zero.

Table 10: Estimation Results of the Diploma Equation

	Without unobserved heterogeneity (1)			With time-invariant unobserved heterogeneity (2)			With time-invariant and time-variant unobserved heterogeneity (3)		
	Coeff.		Std. Err.	Coeff.		Std. Err.	Coeff.		Std. Err.
Exogenous variables									
Female	0.724	***	0.182	2.023	***	0.642	2.794	***	0.858
Cohort 1980	0.034		0.173	0.027		0.345	-0.652		0.456
Calendar day of birth/100	0.166	**	0.084	0.115		0.169	0.018		0.210
Father's education/10	0.079		0.340	1.087		0.764	1.528		0.996
Mother's education/10	-0.144		0.353	1.134	*	0.684	2.007	**	0.943
Number of siblings – Reference: No siblings									
1 sibling	0.162		0.253	-0.140		0.590	-0.639		0.810
2 siblings	-0.003		0.284	-0.723		0.659	-1.509		0.978
3 or more	-0.263		0.311	-1.323	*	0.737	-2.348	**	1.159
Endogenous variables†									
Track – Reference: BSO									
ASO+	1.420	***	0.385	0.787		0.818	-0.171		1.301
ASO−	0.873	***	0.283	0.334		0.659	-0.151		1.126
TSO+	-0.139		0.313	0.712		0.649	-2.285	*	1.318
TSO−	0.231		0.279	0.611		0.687	-0.433		1.076
Total years of delay	-0.441	***	0.138	0.150		0.314	0.647		0.542
Lagged retention	0.042		0.431	3.241	***	1.103	7.170		6.588
Unobserved heterogeneity loading factor									
Loading factor	–		–	-9.422	**	3.999	-23.152	*	14.000

† As students are not allowed to change track at the beginning of the last grade, there is no control for lagged downgrade decision in the diploma equation.

[14] As mentioned before, the last grade of ASO/TSO tracks is the 6th grade, while the last grade of the BSO track is the 7th grade.

6 Conclusions

We empirically analysed the educational choices and performances faced by pupils and/or their parents in the secondary school system of Flanders, the Dutch speaking region of Belgium. In Flanders, tracks are hierarchically ordered and track mobility is permitted only in case of track downgrades. Low achieving students can be asked to resit the grade. We exploited econometric modelling tools and identification analysis to examine the interrelated dynamics of secondary school grade retention, track choices and achievements of a sample of Belgian pupils living in Flanders. We also shed light on the role played by family background and unobserved abilities, especially looking at how unobserved abilities interact with retention episodes in determining schooling pathways.

The empirical analysis was based on the rich schooling information contained in the SONAR dataset, a retrospective survey conducted in Flanders on the 1976, 1978 and 1980 cohorts. Our sample was made up of 4,214 students belonging to the 1978 and 1980 cohorts. We exploit the ample information on secondary school performances and choices to estimate dynamic qualitative choice models.

Concerning the impact of downgrading, we find that pupils downgrading track are more likely to get good evaluations in the next academic year. Although there might be negative effects induced by changing peers and sometimes school, students who come from a higher track are likely to have an excess of knowledge relatively to the new track-level requirements, so they succeed more easily. Moreover, students who have experienced a track change are less likely to downgrade in the following year, meaning that downgrading stabilises the track pathways of students. In addition, and in line with theories in the sociology of education which suggest that educational choices are influenced by social status maintenance and structural risk aversion, we also find that initial track choices and pupils' subsequent track mobility are strongly associated with parental background. Pupils with lower educated parents are more likely to choose vocational tracks and are more like to experience track downgrades.

In contrast to most of previous findings, we find that grade retention has a positive impact on the next evaluation and can permanently affect subsequent performances. The direction of the permanent effect depends on unobserved heterogeneity. While more able students are permanently penalised by retention, less able students benefit from it. The fact that more able students might be permanently penalised by an episode of grade retention suggests that psychological costs dominate possible benefits. For less able pupils the psychological costs might instead be dominated by the benefit of having more time to develop the knowledge and emotional maturity required at each educational grade. We conclude that when assessing the effectiveness of grade retention, one should carefully consider the heterogeneity in responses to grade retention by unobservable behavioural and cognitive abilities.

References

Aakvik, A., K.G. Salvanes, and K. Vaage, "Measuring Heterogeneity in the Returns to Education Using an Education Reform,," *European Economic Review*, 2010, *54*, 483–500.

Alexander, K.L., D.R. Entwisle, and S.L. Dauber, *On the success of failure*, New York: Cambridge University Press, 1994.

Altwicker-Hámori, S. and J. Köllő, "Whose children gain from starting school later? – Evidence from Hungary," *Educational Research and Evaluation*, 2012, *18* (5), 459–488.

Bedard, K. and E. Dhuey, "The persistence of early childhood maturity: International evidence of long-run age effects," *The Quarterly Journal of Economics*, 2006, *121* (4), 1437–1472.

Belot, M. and V. Vandenberghe, "Evaluating the "Threat" Effects of Grade Repetition. Exploiting the 2001 Reform by the French-Speaking Community of Belgium," IRES Discussion Papers 2011-26, Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES) 2011.

Boudon, R., *Education, opportunity, and social inequality: Changing prospects in Western society*, New York: Wiley-Interscience, 1974.

Bratti, M., L. Cappellari, O. Groh-Samberg, and H. Lohmann, "School tracking and intergenerational transmission," in J. Ermisch, M. Jäntti, and T.M. Smeeding, eds., *From parents to children. The intergenerational transmission of advantage*, New York: Russell Sage Foundation, 2012, chapter 13, pp. 311–344.

Breen, R. and J.H. Goldthorpe, "Explaining educational differentials: Towards a formal rational action theory," *Rationality and Society*, 1997, *9* (3), 275–305.

Browman, L.J., "Grade retention: Is it a help or hindrance to student academic success?," *Preventing School Failure*, 2005, *49* (3), 42–46.

Byrd, R.S., M. Weitzman, and P. Auinger, "Oncreased behavior problems associated with delayed school entry and delayed school progress," *Pediatrics*, 1997, *100* (4), 654–661.

D'Haultfœuille, X., "A new instrumental method for dealing with endogenous selection," *Journal of Econometrics*, 2010, *154* (1), 1–15.

Dustmann, C., "Parental background, secondary school track choice, and wages," *Oxford Economic Papers*, 2004, *56* (2), 209–230.

Eide, E.R. and D.D. Goldhaber, "Grade retention: What are the costs and benefits?," *Journal of Education Finance*, 2005, *31* (2), 195–214.

_ **and M.H. Showalter**, "The effect of grade retention on educational and labor market outcomes," *Economics of Education Review*, 2001, *20* (6), 563–576.

Falter, J., G. Ferro-Luzzi, and F. Sbergami, "The effect of parental background on track choices and wages," *Swiss Journal of Economics and Statistics*, 2011, *147* (2), 157–180.

Fredriksson, P. and B. Öckert, "Is early learning really more productive? The effect of school starting age on school and labor market performance," Working Paper 2006:12, IFAU - Institute for Evaluation of Labour Market and Education Policy 2006.

Fruehwirth, J.C., S. Navarro, and Y. Takahashi, "How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects," Working Paper 2011-7, University of Western Ontario, CIBC Centre for Human Capital and Productivity 2011.

Gamoran, A., "Measuring curriculum differentiation," *American Journal of Education*, 1989, *97* (2), 129–143.

—, M. Nystrand, M. Berens, and P.C. LePore, "An organizational analysis of the effects of ability grouping," *American Educational Research Journal*, 1995, *32* (4), 687–715.

Gaure, S., K. Røed, and T. Zhang, "Time and causality: A Monte Carlo assessment of the timing-of-events approach," *Journal of Econometrics*, 2007, *141* (2), 1159–1195.

Goldthorpe, J.H., "Class analysis and the reorientation of class theory: The case of persisting differentials in educational attainment," *British Journal of Sociology*, 1996, *47* (3), 481–505.

Greene, J.P. and M.A. Winters, "Revisiting grade retention: An evaluation of Florida's test-based promotion policy," *Education Finance and Policy*, 2007, *2* (4), 319–340.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales, "Culture, gender, and math," *Science*, 2008, *320* (5880), 1164–1165.

Guyon, N., E. Maurin, and S. McNally, "The Effect of Tracking Students by Ability into Different Schools: A Natural Experiment," *Journal of Human Resources*, 2012, *47*, 684–721.

Hall, C., "The effects of reducing tracking in upper secondary school," *Journal of Human Resources*, 2012, *47* (1), 237–269.

Hallinan, M.T., "Tracking: From theory to practice," *Sociology of Education*, 1994, *67* (2), 79–84.

Hámori, S., "The effect of school starting age on academic performance in Hungary," 2007. Budapest Working Papers on the Labour Market No. 2007/2, Corvinus University of Budapest.

Hartlaub, V. and T. Schneider, "Educational choice and risk aversion: How important is structural vs. individual risk aversion?," SOEP Paper on Multidisciplinary Panel Data Research 433-2012, DIW Berlin 2012.

Haveman, R.H. and B.L. Wolfe, "The determinants of children's attainments: A review of methods and findings," *Journal of Economic Literature*, 1995, *33* (4), 1829–1878.

Heckman, J.J., "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process," in C. Manski and D. McFadden, eds., *Structural analysis of discrete data with econometric applications*, Cambridge: MIT Press, 1981, chapter 4, pp. 179–195.

— **and B. Singer**, "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 1984, *52* (2), 271–320.

Jacob, B.A. and L. Lefgren, "Remedial education and student achievement: A regressor-discontinuity analysis," *Review of Economics and Statistics*, 2004, *86* (1), 226–244.

— **and** — , "The effect of grade retention on high school completion," *American Economic Journal: Applied Economics*, 2009, *1* (3), 33–58.

Jimerson, S.R., "Meta-analysis of grade retention research: Implications for practice in the 21st century," *School Psychology Review*, 2001, *30* (3), 420–437.

Kerr, S.P., T. Pekkarinen, and R. Uusitalo, "School tracking and development of cognitive skills," *Journal of Labor Economics*, 2013, *31* (3).

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

McEwan, P.J. and J.S. Shapiro, "The benefits of delayed primary school enrollment: Discontinuity estimates using exact birth dates," *Journal of Human Resources*, 2008, *43* (1).

Meghir, C. and M. Palme, "Educational Reform, Ability, and Family Background," *American Economic Review*, 2005, *95*, 414–424.

Mroz, T. and G. Picone, "A multiple state duration model with endogenous treatment," HEDG Working Paper 11/19, HEDG, c/o Department of Economics, University of York 2011.

Mroz, T.A., "Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome," *Journal of Econometrics*, 1999, *92* (2), 233–274.

OECD, *Learning for tomorrow's world – First results from PISA 2003* 2004.

Piopiunik, M., "The Effects of Early Tracking on Student Performance: Evidence from a School Reform in Bavaria," Ifo Working Paper 153, Ifo Institute 2013.

SONAR, *Jongeren in transitie* De arbeidsmarkt in Vlaanderen, Jaarreeks 2000, deel 4, Leuven/Brussel: Steunpunt WAV, 2000.

Strøm, B., "Student achievement and birthday effects," 2004. Paper presented at the CESifo-Harvard University/PEPG Conference on "Schooling and Human Capital in the Global Economy: Revisiting the Equity-Efficiency Quandary", CESifo Conference Center, Munich.

Thurow, L.C., "Education and economic inequality," *Public Interest*, 1972, *28* (Summer), 66–81.

Van de gaer, E., H. Pustjens, J. Van Damme, and A. De Munter, "Tracking and the effects of school-related attitudes on the language achievement of boys and girls," *British Journal of Sociology of Education*, 2006, *27* (3), 293–309.

Van Houtte, M., "Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff," *American Journal of Education*, 2004, *110* (4), 354–388.

— , **J. Demanet, and P.A.J. Stevens**, “Self-esteem of academic and vocational students: Does within-school tracking sharpen the difference?,” *Acta Sociologica*, 2012, *55* (1), 73–89.

Xia, N. and N. Kirby, “Retaining students in grade. A literature review of the effects of retention on students’ academic and nonacademic outcomes,” Technical Report, NBER 2009.