Dropout from secondary education: All's well that begins well

Kristof De Witte

TIER WORKING PAPER SERIES
TIER WP 10/01



Dropout from secondary education: All's well that begins well*

Kristof De Witte

Top Institute for Evidenced Based Education Research

Maastricht University

Kapoenstraat 2, 6200 ML Maastricht

k.dewitte@maastrichtuniversity.nl

www.personeel.unimaas.nl/k-dewitte

March 9, 2010

Preliminary version.

Abstract

Despite the increased attention to students leaving secondary education without a diploma, numerous students dropout yearly. This paper makes a distinction between the 'individual perspective' and the 'institutional perspective' of dropping out. The individual perspective considers the probability of an individual student to drop out. It is explored by multinominal logit models, with and without accounting for unobserved heterogeneity. We observe that particularly motivation of the student and interest in schooling of his/her parents are crucial predictors of the individual dropout decision. The institutional perspective focusses on contextual factors and is examined by ordered logit models, both with and without accounting for unobserved heterogeneity. In particular, we discuss the influence of the first year of secondary education by analysing the large differences in the number of dropouts in Dutch first year classes. We observe that, more than motivation, conditions in the first year of secondary education are crucial in shaping the dropping out decision.

Keywords: Dropout decision, Secondary education, Motivation, Logit, Unobserved heterogeneity

JEL-classification: I21, C35

*This paper benefited from discussions during various meetings at the Dutch Ministry of Education. I am particularly grateful to Emily Brounts, Marton Csillag, Martijn Groenleer, Wim Groot, Geert Thyssen, Henriëtte Maassen van den Brink, Marijn Verschelde, Fred Voncken, Kasper van Weekborg, Liesbeth van Welie and participants at the TIER workshop at Groningen University.

1 Introduction

The decision of the European Council to half by 2010 the number of students withdrawing from school without a higher secondary degree, significantly increased the attention of both practitioners and academics to school dropout. There is indeed a sense of urgency as the literature rigorously shows that students dropping out at secondary education have only bleak prospects (Pascarella and Terenzini, 1991). The consequences of school dropouts are situated at three levels. Firstly, there are private costs such as, in comparison to non-dropouts, higher unemployment risks (e.g., Psacharopoulos and Layard, 1979), lower health status (e.g., Groot and Maassen van den Brink, 2007) or lower educated children (e.g., Bowles, 1972). Secondly, there are increased costs to society by increased risk for criminal activities (e.g., Lochner and Moretti, 2004), lower social cohesion (e.g., Milligan *et al.*, 2004) or lower rate of economic growth (Hanushek and Wößmann, 2007). Finally, there are fiscal consequences due to lower tax revenues, higher unemployment allowances or higher health costs (for an overview, Psacharopoulos, 2007).

Following the European Council, we define a dropout student as a young person between the age of 12 and 23 who leaves education without a degree or with only a lower-secondary level degree (European Commission, 2006). In this sense, the definition represents an output-oriented indicator (i.e., a simple head count indicator) and does not reflect the cognitive skills of pupils. As such, a graduating student with very low cognitive skills will not be considered as a dropout (Psacharopoulos, 2007).

Examining the individual dropout decision boils down to estimating those influences which significantly increase the probability that a student will leave secondary education without a diploma. However, as an outsider, who tries to obtain insights in the dropout decision of the average student by mimicking his/her way of thinking (i.e., by selecting a broad set of both exogenous and motivational factors), we should be aware that, in the end, it is the individual student who decides to leave or to stay in school. As such, even the broadest specified model will suffer from a large noise term which captures the unobserved heterogeneity among the individuals in the sample. We explicitly account for unobserved heterogeneity in the estimated multinominal and ordered logit models. Presenting the outcomes for both estimation assumptions allows us to obtain insights in the size of the unobserved heterogeneity bias.

Similar as in Rumberger (2001), we make a clear distinction between the 'individual perspective' of dropping out and the 'institutional perspective'. Whereas the former focusses on the factors which shape the individual decision to drop out (i.e., individual background and motivation), the latter focuses on the contextual factors (i.e., characteristics of the class and peers) which might influence the dropout decision. The contribution of this paper arises from examining the background characteristics which constitute the individual and institutional

perspective.

Firstly, within the individual perspective, we mimic as well as possible the way of thinking of the individual student. These thoughts create heterogeneity among students as they differ because of their own exogenous characteristics (e.g., gender, ethnicity, age, etc.), the background of their parents (e.g., schooling of the parents, interest in schooling, etc.), and the influence of the schooling environment (e.g., rural versus urban, school track, etc.). Each of those deterministic background characteristics shape the individual motivation of the student. We capture motivation by, e.g., the student's opinion about the teachers, the opinion of teachers about the student, their attention during classes, the number of times they are truant or their school career. To empirically test the individual decision to dropout, we use an exceptionally rich and large sample on the Dutch secondary education (in particular Voortgezet Onderwijs Cohort Leerlingen; VOCL).

Secondly, based on results for post-secondary education, we pay within the institutional perspective special attention to the conditions in the first year of secondary education. In particular, we observe that for some first year secondary education classes none of the students is (later) dropping out, whereas in other classes up to seven students drop out. As the literature on post-secondary education (e.g., Tinto, 1975; Pascarella and Terenzini, 1980; Bowlby and McMullen, 2002; Jansen and Bruinsma, 2007) and on middle education (Alexander *et al.*, 1997; Garnier *et al.*, 1997) shows, this is not a coincidence: the first year of (secondary) education is crucial. In the fourth section of the paper, we examine the probability that a student belongs to a class with many dropout students. As such, we determine the conditions which go along with making good first year classes (i.e., classes with few dropouts) and bad first year classes (i.e., with higher dropout levels). To do so, we focus on class characteristics such as number of students in the class, average ability, standard deviation of the abilities, percentage of boys, or ethnicity at school level.[1] It is remarkable to observe that, while the dropout decision at the individual level is largely driven by motivational factors, at a class level individual motivation becomes insignificant while class determinants are important. In this sense, our results are close to the observation of Alexander *et al.* (1997, p.98): "*In the present instance, all these themes come into play long before anyone is thinking about dropout per se. We are not saying that what happens in first grade necessarily seals children's fates, but prospects for "re-engagement" later are not good when children are plagued early in their school careers*".

The remainder of the paper is structured as follows. Section 2 briefly reviews the data at hand and discusses some conceptual models. In Section 3 we use a multinominal logit model, with and without unobserved heterogeneity, to examine the drivers of the individual decision

[1] Note that, due to data limitations, we do not attempt to solve the causality issue. However, as the schools have no *a priori* information on the motivation of the students, we can argue that conditions in the first year of secundary education determine dropout probability and not the other way around.

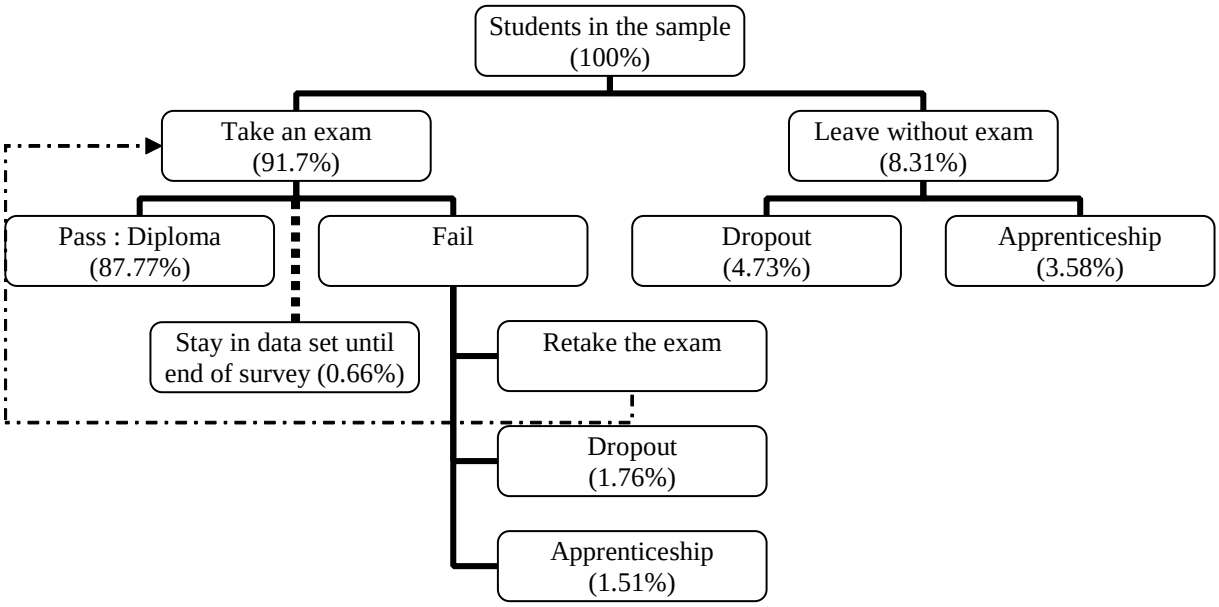

Figure 1: Systematic presentation of dropout students (percentage of sample population in brackets)

to drop out. On the contrary, Section 4 is concerned about the collective 'peer' decision. We examine whether conditions (where class characteristics serve as a proxy) in the first year secondary education influence the dropout. In a final section we offer our (policy) conclusions.

2 How to look at dropping out?

2.1 The data

To examine the dropout decision of students in secondary education, we use the Dutch VOCL data (Voortgezet Onderwijs Cohort Leerlingen). This paper does not attempt to describe the Dutch school system (see Tieben and Wolbers, 2008; Dodde and Leune, 1995), but only takes advantage of the rich VOCL data set. The VOCL data, collected by Statistics Netherlands and financed by the government, follow a cohort of pupils from the first year of secondary education until they leave school (either with or without a degree). The data consist of a representative sample of 20,331 students in 330 schools and contain various questions on (1) the class and school type the pupil is taking, (2) the educational attainments (e.g., test scores), (3) attitude of the pupils (e.g., attention to homework, extra-curriculum activities, attitude towards schooling), (4) attitude of parents (e.g., interest in schooling, who takes which decision within the household) and (5) socio-economic variables (e.g., country of birth,

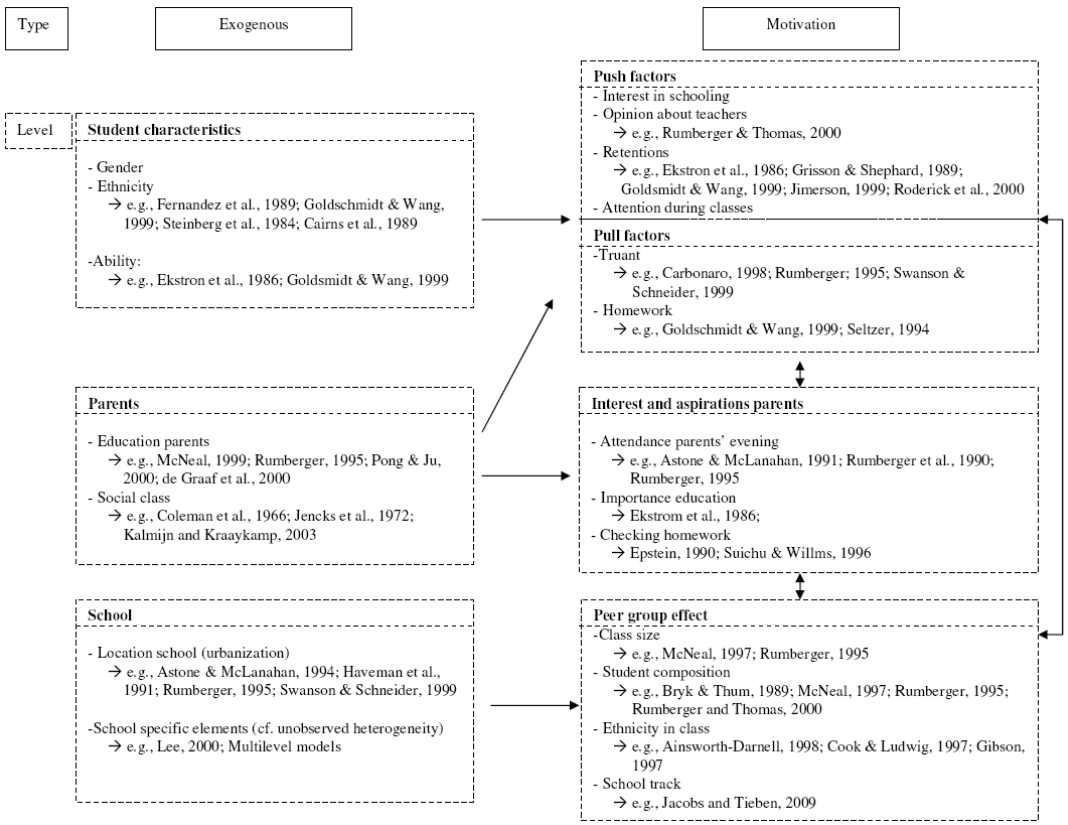

Figure 2: Potential drivers for dropout

education of parents, socio-economic status).[2] The data are collected immediately from the schools, and from surveys from pupils and parents in the beginning of the first and third cohort year. In the current study, we use variables which are gathered in the first cohort year (unless differently noted). Currently, the VOCL is at its third round (although the last cohort, started in 1999, is still ongoing). In the current study, we use in particular the second round, the 1993 cohort (as this is the most recent, fully completed round).

The cohort starts in the first year of secondary education and follows the students throughout their academic career. This is systematically presented in Figure 1. Students who died, had a protracted illness or moved abroad were removed from the sample.[3] As such, we obtain a sample of 17,697 students. The paper at hand assumes that every student has a choice option of (1) obtaining a diploma, (2) dropping out, (3) entering an apprenticeship program and (4) staying at school (due to repeating classes). The choice options are often restricted by conditions beyond the control of the individual (e.g., the apprenticeship program is only available for vocational education students). On top of that, the actual decision (within the feasible option) is determined by both individual (e.g., ability) and background (e.g., parents) conditions. In Section 4 we explore this interplay with the conditions of the first year of secondary education.

Consider the four groups of students. Firstly, about 88% of the students in the sample take and pass the end exam. We observe from Table 1 that the majority of students with a diploma leave school between the age of 16 and 20. This is not surprising as compulsory education was until the age of 16 (since 2003, this changed to age of 18). Therefore, students could leave school at the age of 16 with a diploma. Remark that some students are younger than the other students in their class, and as such, can take the end exam at the age of 15.

A second group of students leave school without a secondary education diploma and are defined as dropouts. As is visualized in Figure 1, those students left school either without taking an exam or after (frequently) failing the exam. In total, there are 1,148 dropout students (6.49%). As presented in Table 1, the majority of the dropouts leave education between the third and the sixth cohort year, between the normal age of 15 (third cohort year) and 18 (sixth cohort year). Again, given the age limit of compulsary education, this is not surprising.

Thirdly, we consider students who leave regular education to start a 'learning on the job' program. In our sample, 5.09% of the students enter an apprenticeship program, from which

[2] Although students cannot refuse to be included in the VOCL sample, they can refuse to fill in the student survey. This could potentially create a selection bias. However, only very few students refused to fill in the survey (less than 1.5% of the 20,000 students refused during the 8 cohort years), such that the sample is representative.

[3] Students who moved to schools not belonging to the original VOCL sample were further followed on these schools.

Table 1: Age of students at the event time

Age	Diploma		Dropout		Apprentice		Stay in school		Total	
	Number	%	Number	%	Number	%	Number	%	Number	%
12	0	0%	10	1%	2	0%	0	0%	12	0%
13	0	0%	32	3%	9	1%	0	0%	41	0%
14	1	0%	95	8%	47	5%	0	0%	143	1%
15	1,022	7%	226	20%	232	26%	0	0%	1,480	8%
16	2,303	15%	335	29%	276	31%	0	0%	2,914	16%
17	2,180	14%	251	22%	203	23%	0	0%	2,634	15%
18	1,938	12%	126	11%	104	12%	0	0%	2,168	12%
19	5,589	36%	58	5%	22	2%	80	69%	5,749	32%
20	2,341	15%	14	1%	5	1%	34	29%	2,394	14%
21	150	1%	1	0%	0	0%	2	2%	153	1%
22	7	0%	0	0%	1	0%	0	0%	8	0%
23	1	0%	0	0%	0	0%	0	0%	1	0%
Total	15,532	100%	1,148	100%	901	100%	116	100%	17,697	100%

1.51% after one or several failed exams. The typical age to start an apprenticeship is similar to the age of dropping out: between 15 and 18 years old. As the apprenticeship program is only a choice option for students in vocational training, we control in the analysis for learning subject.

Finally, there is a minority of students (0.66% of the sample population) who are still enrolled at school by the end of the cohort study. As it can be interesting to examine why those students did not dropout, we include them in the analysis (robustness analysis indicates that removing these 116 students from the sample, does not significantly change our results).

2.2 Conceptual dropout model: A brief review

In the examination of the dropout decision of students, the (academic) literature is dramatically suffering from unobserved heterogeneity (DesJardins *et al.*, 1999). Indeed, if asked to individual students about their reasons for dropping out, they always point to a mixture of causes (e.g., Bowlby and McMullen, 2002; Bridgeland *et al.*, 2006). In this paper, we do not attempt to fully review the extensive literature on dropout (for an excellent review see Rumberger, 2001), but we combine two conceptual models in our analysis. On the one hand, started from Spady (1970) and Tinto (1975), the student integration model focusses on school related factors (i.e., institutional factors) such as motivation, ability and school characteristics. On the other hand, initiated by Bean (1978), the student attrition model discusses the importance of non-institutional influences outside the school environment. We believe that it is a combination of institutional and non-institutional factors that shape the decision to dropout. In an attempt to combine the student integration and student attrition model, we suggest a vertical and horizontal classification (see Figure 2).

The vertical classification corresponds to the difference between the exogenous background characteristics and the motivational influences, what is labeled by Roemer (1998) as 'circumstances' versus 'effort'. While the exogenous background characteristics are deterministic to

the students, the motivation is shaped and influenced by the exogenous environment. The horizontal classification makes a distinction between influences arising from the students, the parents and the school. The vertical and horizontal dimension are undoubtly related. Firstly, background characteristics of the students result in motivational factors which we classify as push and pull factors (following Rumberger, 1987). Push factors arrive from the absence in motivation to stay at school and 'push' the student away from school. In a different nuance (and obviously closely related), pull factors originate from the school environment and 'pull' the students away from school.[4] Although the literature suggests that student background characteristics as, e.g., gender have a significant effect on dropout, it is arguable that the dropout decision does not originate from the gender (i.e., being male or female) but rather from the student's motivation (which is often correlated to students' gender). If in a multivariate analysis, where the researcher controls for other influences, gender turns out to have a significant effect on the individual dropout decision, the effect is probably driven by unobserved heterogeneity in gender.[5] Secondly, education and social class of the parents influence the parents' interest and aspirations. Similar to the student's characteristics, it can be argued that the social class itself is of minor importance in comparison to the motivation of the parents. Finally, school characteristics as location shape the peer group effects. To obtain some insights in the data, we present them bivariately in Table 6 in Appendix.

3 Blaming the environment? - A multinominal logit

3.1 Multinominal logit

To analyse the magnitude and direction of the influence of explanatory variables on the decision to leave school with or without a diploma, or to enter the apprenticeship program, we estimate a multinominal logit (Mlogit) model (McFadden, 1973). The Mlogit model is used to estimate relationships between (ordered or unordered) polytomous dependent variables and multivariate explanatory variables. In particular, it estimates the probability of an event occurring. As such, it is a straightforward extension of the logistic model, which estimates the model by maximum likelihood after transforming the dependent variable into a logit variable (i.e., the log of the odds of the dependent occurring or not) (e.g., Train, 2003).

Suppose there are J ($k = 1, \ldots, j, \ldots, J$) different events which may be chosen by individual i. Given the (exogenous) characteristics of individual i (X_i), the probability of selecting event j equals:

$$\Pr(y_i = j \mid X_i) = \frac{\exp(X_i \beta_j)}{\sum_{k=1}^{J} \exp(X_i \beta_k)} \tag{1}$$

[4] As we assume that both the push and the pull factors are student-initiated, our distinction is different from the 'voluntary withdrawal' and 'involuntary withdrawal' of Rumberger (2001).

[5] In the multivariate analysis, we run logit models with and without controlling for unobserved heterogeneity at school level.

where β denotes the unknown (to be estimated by maximum likelihood) parameter which presents the log odds of being in the target group relative of being in the reference group. However, this set of equations will be unidentified (i.e., there are infinite solutions). Therefore, we make one of the categories a reference category (say, category 1) and set its coefficients equal to 0. This brings us to the following model:

$$\Pr(y_i = 1 \mid X_i) = \frac{1}{1 + \sum_{k=2}^{J} \exp(X_i \beta_k)} \tag{2}$$

$$\Pr(y_i = j \mid X_i) = \frac{\exp(X_i \beta_j)}{1 + \sum_{k=2}^{J} \exp(X_i \beta_k)}. \tag{3}$$

In the remainder we will express the estimated coefficients immediately in relative risk ratios, i.e., in the $\exp(\beta_k)$ rather than β_k. A relative risk of $\exp(\beta_k) = 1$ denotes that there is no difference between the reference group and the evaluated group. A relative risk larger (smaller) than 1 indicates that the event will occur more (respectively, less) likely in the evaluated category than in the reference category.

In our application, we assume students with a diploma as reference category. As such, we can compare the probability of belonging to another group (in casu: dropout, apprenticeship or staying) with the probability of belonging to the reference group (i.e., leaving school with a diploma).

3.2 Multinominal logit - Results

The results of the multinominal logit regression are presented in Table 2. It is interesting to observe that, once controlled for motivation, gender does not have a significant effect on belonging to one or another group. In other words, given the background and motivation of the students, males do not dropout more frequently than females.[6] This contrasts to previous work of Fernandez *et al.* (1989), Goldschmidt and Wang (1999) and Rumberger (2001). We can make a similar observation for Moroccan students, truants and work of the parents. Concerning the latter, it is notable and intuitively explainable that students whose parents work in one-man businesses or are self-employed enter significantly more frequently an apprenticeship program. Increasing abilities of students (as proxied by the average of the cito-standarized test scores for math, languages and information; cito-test is a national and standardized test at the end of the primary school which is used as an advisory tool for the school track) initially reduces the probability of dropping out (entering an apprenticeship program, respectively), however, as revealed by the squared term, this impact fades out for smarter students. As a last exogenous student characteristic, students who are older at the

[6] Although in a multivariate analysis we control for the mutual influence of the variables, due to a potential correlation between the unobserved characteristics (captured in the error term) and the observed characteristics, the estimation outcomes could be biased. An instrumental variable approach may be appropriate, yet due to data limitations the proper instruments are not available.

start of the cohort drop out significantly more than younger students (similar for entering the apprenticeship program).

All push factors have a significant effect on dropout. Students who like school, have a favorable opinion about their teachers in general and about their math teacher in particular, have a normal school career, have teachers who are pleased with their results and pay attention during classes have a lower probability of dropping out. The pull factors point out that the longer it takes before a student is starting his/her homework, the higher the dropout probability is.

Besides the characteristics of the student, also the opinion and background of the parents is important. More highly educated parents, who attend parents' evenings, talk about school at home and think that a high degree is important (i.e., educational aspirations), have a lower probability that their children will drop out (cf. Kalmijn and Kraaykamp, 2003). Indeed, it has been argued by, among others, de Graaf *et al.* (2000) that children from highly educated parents receive cultural resources as knowledge, tastes and preferences which favor the educational career. Inversely, extensive parental control on homework and extensive affiliation with homework, the higher the probability is that their child will drop out. It seems that parental care is important, however, students should make their homework themselves.[7] Neighbourhood significantly affects the dropout decision as students are dropping out more frequently in more urban areas. Finally, our results confirm the trend which is observed by de Graaf and Ganzeboom (1993) in that occupational status is becoming less important (they examined the period 1891-1960). In this 1993 wave of data, the occupational status of the parents does not influence the dropout decision. Nevertheless, the literature shows that effective programs to reduce dropout in high schools are often a collaboration with the school and the parents (e.g., van Heusden Hale, 2000; Gandara *et al.*, 1998).

In sum, with regard to students' individual decision to drop out of school, the estimation results in Table 2 reveal that particularly motivation is crucial. In this sense, our results follow the conclusions of Dekkers and Driessen (1997) who observe that students decide to dropout on the basis of emotional aspects rather than performance aspects. Therefore, it can be interesting to examine why students become unmotivated. In the remainder of the paper, we explore two potential explanations. Firstly, although we tried to control for a broad set of exogenous variables (in an attempt to mimic a student's way of thinking), as a rule, there is still unobserved heterogeneity among students because we cannot explicitly include all determinants in the analysis. In the next subsection, we explicitly control for this unobserved heterogeneity.[8] Secondly, and following the literature in post-secondary dropping

[7] It is likely that extensive parental control is a consequence of weak performance at school, and not a cause for drop out *an se*.

[8] We insist on presenting the results without accounting for unobserved heterogeneity as (1) the model without unobserved heterogeneity is common practice in the literature, and (2) the comparison allows the reader to obtain insights on the extent to which the results differ.

out, in Section 4 we explore the influence of first year high school characteristics on the (later) dropout decision.

3.3 Multinominal logit with unobserved heterogeneity

To model a multinominal logit with unobserved heterogeneity (also known as variance of the disturbances), we add unobserved individual effects α_i to the probability in equation (1):

$$\Pr(y_i = j \mid X_i, \alpha_i) = \frac{\exp(\alpha_{ij} + X_i\beta_j)}{\sum_{k=1}^{J} \exp(\alpha_{ik} + X_i\beta_k)}. \tag{4}$$

Similar as before, the unknown coefficients (α_i and β_j) are estimated by maximum likelihood. In contrast to the situation without unobserved heterogeneity, we now have to estimate the likelihood by integrating over the distribution of the unobserved heterogeneity (see, e.g., Haan and Uhlendorff, 2006). Due to the unobserved character of α_i, there is no analytical solution to this integration such that an approximation is needed. As the adaptive quadrature approximation (i.e., a proxy of the integral by a specified number of discrete points after which a distribution of the unobserved heterogeneity is made) is the most accurate way to approximate the unknown integral, we use this procedure.

If in the multinominal logit model we account for unobserved heterogeneity among schools (and, as such, acknowledge that we cannot fully capture the students' way of thinking) we observe from Table 3 that the estimations do not dramatically change. In contrast, variables that have a significant influence on the decision in the traditional multinomial logit, do still have a significant influence when allowing for unobserved heterogeneity. Also the direction of the effects is similar. Additionally to the traditional multinominal logit estimations from Table 2, we observe that females choose less to dropout, as do children from middle employees.

The strong similarities between the estimations give us confidence on the estimations. This confirms the conclusion that in the individual decision to dropout particularly motivational factors are important. This motivation should come from both the students and their parents, although the latter should not exaggerate: a too strict control on homework inhibits the students so that, in the end, they have a higher probability of dropping out.

4 Blaming the school? - An ordered logit

Section 3 has suggested, by using a multinominal logit, that motivation is a crucial factor in the individual choice of students to drop out. As in the current debate dropout students (those leaving school without a diploma) obtain the highest attention, in the remainder of the paper we exclusively focus on this group. At first sight, the importance of motivation seems frustrating for schools, teachers and policy makers as motivation is often intrinsic, coming from the immediate environment and, as such, difficult to influence. Nevertheless, there might be a large influenceable factor at class level which arises from three issues. A

Table 2: Multinominal logit regression

	Dropout		Apprenticeship		Stay in Sample	
A. Characteristics students						
A1. Exogenous characteristics students						
Gender (male = ref)	0.8889		1.0239		0.8408	
	(-0.1209)		(-0.765)		(-0.4069)	
Homeland mother (NE = ref)						
Morocco	1.1592		1.0307		4.8803	**
	(-0.4681)		(-0.9038)		(-0.013)	
Surinam / Antilles	1.4282	*	1.354		2.0801	
	(-0.0712)		(-0.1604)		(-0.2324)	
Turkey	1.5603	**	1.1217		1.2407	
	(-0.0106)		(-0.6069)		(-0.8354)	
Other	2.0633	***	1.3998	**	1.5022	
	(0.0001)		(-0.0154)		(-0.3119)	
Ability	0.8945	***	0.9082	***	1.0775	
	(0.0001)		(0.0001)		(-0.2917)	
Ability * ability	1.0014	***	1.0014	***	0.9997	
	(0.0001)		(0.0001)		(-0.7556)	
Age	1.3221	***	1.3763	***	0.7031	
	(-0.0002)		(-0.0001)		(-0.1303)	
A2. Ideas and behavior students						
A21. Push factors						
"I like school"	0.8533	***	0.9494		0.9181	
	(-0.0001)		(-0.2320)		(-0.4931)	
"Opinion about math	0.7479	***	0.8776		0.7377	
teacher"	(-0.0007)		(-0.1328)		(-0.1247)	
"Teachers do their	0.8247	***	0.7568	***	1.0986	
best at school"	(-0.0018)		(0.0001)		(-0.6381)	
Number of retentions	0.7261	***	0.8556		1.4199	
	(-0.0005)		(-0.1242)		(-0.3322)	
"Teachers are pleased	0.8034	***	0.8665	**	0.8829	
with my results"	(-0.0001)		(-0.0174)		-0.4916	
"I pay attention during	0.8773	**	0.938		1.3332	*
explantions"	(-0.0225)		(-0.2964)		(-0.0983)	
A22. Pull factors						
Truant	0.9107		0.9609		1.1149	
	(-0.2883)		(-0.6428)		(-0.5808)	
Start homework after	1.1144	***	1.1124	***	0.9055	
	(0.0001)		(0.0001)		(-0.3119)	
B. Characteristics environment						
B1. Exogenous characteristics parents						
Education parents	0.8572	***	1.0295		0.8539	
	(-0.0021)		(-0.5577)		(-0.2125)	
Work (worker = ref)						
One-man business	1.1619		1.4407	**	1.507	
	(-0.2736)		(-0.0116)		(-0.3713)	
Self-employed	1.0859		1.2207	*	1.6948	
	(-0.3737)		(-0.0549)		(-0.1008)	
Lower employee	0.9391		0.8203		1.5713	
	(-0.6838)		(-0.2329)		(-0.258)	
Middle employee	0.7935		0.9615		1.7698	
	(-0.1102)		(-0.7781)		(-0.115)	
Higher employee	0.8821		1.0761		1.6006	
	(-0.4750)		(-0.6488)		(-0.257)	
B2. Interest in schooling and aspirations parents						
Attended parents' evening	0.3874	***	0.5607	***	1.0921	
	(0.0001)		(0.0001)		(-0.7899)	
"A high degree is important"	0.9115	*	-1.0303		0.8301	
	(-0.0966)		(-0.604)		(-0.1667)	
"Affiliation with homework"	1.1964	***	1.2299	***	1.258	
	(-0.006)		(-0.0017)		(-0.1673)	
"Talking about school at home"	0.9124	*	0.9472		1.1686	
	(-0.0927)		(-0.3483)		(-0.3153)	
Checking homework	1.1521	***	1.1208	***	0.8053	
	(-0.0006)		(-0.0099)		(-0.1212)	
B3 Country side	0.8532	***	0.8391	***	0.7960	***
	(0.0000)		(0.0000)		(0.0052)	

χ^2=1.4e3 (p=0.0000); p-values between brackets; ***, **, * denote significance at, respectively, 1, 5 and 10%.

Table 3: Multinominal logit regression with unobserved heterogeneity

	Dropout		Apprenticeship		Stay in Sample	
A. Characteristics students						
A1. Exogenous characteristics students						
Gender (male = ref)	0.8698	*	1.0011		0.8285	
	(0.0729)		(0.9890)		(0.3685)	
Homeland mother (NE = ref)						
Morocco	1.0941		0.9710		4.6343	**
	(0.6673)		(0.9083)		(0.0175)	
Surinam / Antilles	1.3162		1.2421		1.8427	
	(0.1752)		(0.3254)		(0.3215)	
Turkey	1.4736	**	1.0565		1.1487	
	(0.0310)		(0.8091)		(0.8939)	
Other	2.0254	***	1.3746	**	1.4566	
	(0.0000)		(0.0235)		(0.3512)	
Ability	0.8926	***	0.9057	***	1.0728	
	(0.0000)		(0.0000)		(0.3234)	
Ability * ability	1.0014	***	1.0014	***	0.9997	
	(0.0000)		(0.0000)		(0.7503)	
Age	1.3195	***	1.3734	***	0.7048	
	(0.0003)		(0.0001)		(0.1335)	
A2. Ideas and behavior students						
A21. Push factors						
"I like school"	0.8620	***	0.9610		0.9324	
	(0.0002)		(0.3651)		(0.5753)	
"Opinion about math	0.7375	***	0.8678		0.7409	
teacher"	(0.0005)		(0.1094)		(0.1336)	
"Teachers do their	0.8306	***	0.7625	***	1.1027	
best at school"	(0.0030)		(0.0001)		(0.6254)	
Number of retentions	0.7199	***	0.8480		1.3996	
	(0.0004)		(0.1075)		(0.3522)	
"Teachers are pleased	0.8018	***	0.8658	**	0.8900	
with my results"	(0.0001)		(0.0177)		(0.5217)	
"I pay attention during	0.8816	**	0.9422		1.3373	*
explanations"	(0.0293)		(0.3338)		(0.0948)	
A22. Pull factors						
Truant	0.9087		0.9533		1.0935	
	(0.2864)		(0.5866)		(0.6534)	
Start homework after	1.1174	***	1.1141	***	0.9065	
	(0.0000)		(0.0000)		(0.3194)	
B. Characteristics environment						
B1. Exogenous characteristics parents						
Education parents	0.8598	***	1.0257		0.8495	
	(0.0028)		(0.6119)		(0.1973)	
Work (worker = ref)						
One-man business	1.1739		1.4680	***	1.5214	
	(0.2476)		(0.0085)		(0.3620)	
Self-employed	1.1040		1.2385	**	1.7026	*
	(0.2965)		(0.0426)		(0.0991)	
Lower employee	0.9367		0.8240		1.5769	
	(0.6745)		(0.2469)		(0.2557)	
Middle employee	0.7707	*	0.9451		1.7534	
	(0.0748)		(0.6875)		(0.1219)	
Higher employee	0.8666		1.0538		1.5616	
	(0.4203)		(0.7478)		(0.2844)	
B2. Interest in schooling and aspirations parents						
Attended parents' evening	0.3739	***	0.5399	***	1.0507	
	(0.0000)		(0.0000)		(0.8816)	
"A high degree is important"	0.9004	*	1.0182		0.8247	
	(0.0627)		(0.7555)		(0.1536)	
"Affiliation with homework"	1.1983	***	1.2293	***	1.2603	
	(0.0061)		(0.0019)		(0.1646)	
"Talking about school at home"	0.9105	*	0.9468		1.1642	
	(0.0890)		(0.3494)		(0.3291)	
Checking homework	1.1489	***	1.1180	**	0.8018	
	(0.0009)		(0.0124)		(0.1148)	
B3. The neighborhood						
Country side	0.8569	***	0.8467	***	0.8144	**
	(0.0000)		(0.0000)		(0.0133)	
Constant	58.3678	***	1.5144		0.0049	**
	(0.0000)		(0.5296)		(0.0121)	

p-values between brackets; ***, **, * denote significance at, respectively, 1, 5 and 10%.

first issue originates from the extensive literature on post-secondary and middle education which showed that the first year of a new education level is crucial to dropout. Indeed, in the first year of a new education level a heterogeneous group of students is put together in a new school and a new class. The students are unknown to each other, have different backgrounds and experiences, and most importantly, differ in terms of educational knowledge. Moreover, the students, coming from a relatively well protected and structured environment in primary education, are entering a new environment which requires more initiative and independence and is (slightly) less structured. The literature shows (for post-secondary education) that students who are struggling in the first year, have a significant higher probability of dropping out in their later school career (e.g., Tinto, 1975; Pascarella and Terenzini, 1980; Alexander *et al.*, 1997; Garnier *et al.*, 1997; Bowlby and McMullen, 2002; Wilcox *et al.*, 2005; Jansen and Bruinsma, 2007). The arguments that apply for post-secondary education could also go for secondary education; in fact, given the crucial age of the students (12 versus 18 years old) the arguments might even be more pronounced. A second issue arises from the particular application, the Dutch education system. As indicated by Dodde and Leune (1995), the Dutch educational system is typically tracked with a strong selection barrier between primary and secondary education. This barrier makes the first year of secondary education even more crucial. A third reason why class level might be important is to do with peer effects. Carbonaro (1998) showed that a student whose friends drop out has an increased likelihood of dropping out of school himself.

Following previous literature, we assume that the impact of an early stage of secundary education is large. From the data, we observe that some first year classes have only few (or non) dropouts, whereas other classes have a very high (absolute) number of dropouts (note that the students are not necessarily dropping out in the first year). The clustering of dropouts in particular classes is presented in Table 4. We observe that 34% of the students in the first year were in a class in which, in the end, everybody obtained a diploma. For 280 first year classes, only 1 student dropped out (later on). In the remaining 306 classes there was more than one student who left school without a degree; therefore we label these classes 'dropout classes'. As the dropout classes are a typical problem in vocational education, we estimated several robustness checks with only students from vocational training. The outcomes, available upon request, show a high robustness of the results with all observations (and controlling for the subject the student is taking).

To examine the probability of a student to be part of a class with a high drop out propensity, we estimate an ordered logit model which conditions on individual and class characteristics. As an extension of the traditional logit model, the ordered logit estimates ordinal dependent variables. Modelling the number of dropouts in a class as a dependent variable against various explanatory variables allows us to analyse what drives the probability

Table 4: Number of dropouts per class

Number of dropouts in first year class	Number of students in a class with x dropouts	Percentage	Number of classes
0	6,034	34.14	284
1	5,630	31.85	280
2	3,078	17.41	152
3	1,764	9.98	91
4	740	4.19	39
5	271	1.53	16
6	106	0.6	5
7	52	0.29	3

for a student to belong to a class with more dropouts. Before we are explaining our results, we provide a brief methodological background on the ordered logit model.

4.1 Ordered logit

In contrast to Section 3, in this section we consider a discrete independent variable which is ordinal (i.e., the categorical variable is ordered). In particular, we would like to estimate the impact of exogenous influences X_i (i.e., estimate β) on the ordinal y_i (where y_i represents the number of students in the first year of secondary education who (later) dropout of school). Whereas multinominal logit models fail to account for the ordinal nature of y, an ordered logit model is the appropriate technique (Greene, 2003).

To capture the ordinal nature of the response variable, the ordered logit model starts from the cumulative probability, i.e., the odd that an individual i is in category j or higher:

$$\Pr(y_i \leq j \mid X_i) = \sum_{k=1}^{j} \Pr(y_i = k). \tag{5}$$

The ordered logit model is obtained by translating the cumulative probability into a cumulative logit:

$$\begin{aligned} \Pr(y_i \leq j \mid X_i) &= \log \tfrac{\Pr(y_i \leq j)}{1-\Pr(y_i \leq j)} \\ &= \gamma_j - \beta X_i. \end{aligned} \tag{6}$$

From equation (6) we can observe that ordered logit models are proportional odd-models in the sense that the odds ratio of the event is independent of the category j. In other words, we assume that for each of the categories the coefficients β are equal (i.e., an increase in X_i affects the log-odds similarly), while the intercepts γ may differ (i.e., these so-called cutpoints estimate the logit of the odds of being equal or less than category j).

Interpretation of the ordered logit is similar as before, although the dependent variable now counts J groups. An odds ratio (i.e., $\exp(\beta)$) larger than 1 denotes a higher likelihood of belonging to a higher (ordinal) category.

4.2 Ordered logit - results

In the analysis, the first set of explanatory variables correspond to the background and motivational characteristics of previous estimations. Basically, this yields two advantages. On the one hand, it allows us to examine the similarities between the individual decision to drop out and the collective number of dropouts at class level. On the other hand, by including individual and class characteristics in the ordered logit regression, class characteristics are distinguished from individual differences among students. The class characteristics can be divided into three subgroups. Firstly, variables capturing the heterogeneity in abilities as a proxy for student composition (cf. Bryk and Thum, 1989; McNeal, 1997; Rumberger, 1995). This group of variables consist of (1) the mean of student abilities in the class, (2) the standard deviation of abilities, (3) the skewness of abilities, and (4) the track of secondary education (pre-university education (vwo), senior general secondary education (havo), and prevocational secondary education (mavo), or, as this is possible in the first year, a combination of these tracks). Secondly, we include variables capturing heterogeneity in student and parental (exogenous and motivational) characteristics: (1) standard deviation of the age of the students in the class, (2) the percentage of boys in the class, and (3) whether for the parents the school quality was important in determining the school choice. Finally, we proxy some school characteristics as (1) the class size, and (2) the ethnicity at school by including the percentage of Dutch students at school (relative to the total number of students).[9] The results are presented in Table 5.

Looking at the determinants of higher and lower dropout rates yields interesting insights. The exogenous and motivational factors that determined the individual decision to drop out become mostly insignificant when assessing the peer dropout.[10] From the original exogenous variables, only six variables remain significant. The variables can be interpreted as generators of positive or negative externalities as they estimate the odds of belonging to a class with, respectively, less or more dropouts. For example, if the individual student likes school, he/she generates positive externalities to the other students as less students drop out of his/her class (i.e., given that all other variables are held constant, an increase of one unit in 'like school' reduces the odds of being in a higher group by 0.94). Similarly, if the teachers are considered to do well, the number of students dropping out of the class reduces. In a sense, this confirms previous work of Mainhard (2009) who found that the relationship between teacher and student is very stable over time. A truant seems to have a negative externality on his/her fellow students, as does the delay with which students start with homework. Finally, if the

[9] We do not include school size because of potential endogeneity: as parents can choose the school, the as better perceived school will attract more students.

[10] Extensive robustness analyses show that this is not an artifact of the modelling technique (ordered logit versus multinominal logit before), nor from the additional independent variables (we added class characteristics). Indeed, estimating the impact on the individual dropout levels by multinominal logit while adding the class characteristics as explanatory variables, delivers similar (both in significance and in level) results for the exogenous and motivational independent variables. These results are available upon request from the author.

parents are concerned about students' homework, the student significantly ends up more in a class with fewer dropouts (remark that we control for the parental opinion of school quality).

In contrast to the poor explanatory value of students' motivation on the number of dropouts per class, the class characteristics of the first year of secondary education predict the number of future dropouts relatively well.[11] Classes located on the countryside have less future dropouts. Higher average abilities in the class (i.e., on average smarter students) decrease also the number of dropouts. This is in line with Hanushek *et al.* (2003) who found a linear relationship between peer group quality and student performance. Surprisingly, if students more from each other in terms of abilities (i.e., a higher standard deviation) the number of dropouts does not increase, but, in fact decreases. It seems that intelligent students in the class generate positive externalities for the other students. The more the students differ in age, the higher the number of (future) dropouts. Similarly, even controlled for the individual gender and school track of the student, classes with more boys do worse than classes with more girls. This suggests that boys are more impressionable to peer effects. Following the literature, we find that class size has a negative impact on the number of dropouts, as does an increased ethnicity in the school. If, on average, the parents in the class consider the quality of the school as an important factor in determining the school choice, the number of dropouts in the class is lower. Finally, controlled for the ability of the student, school track does not have a significant influence. This confirms previous research (Hanushek and Wößmann, 2006). Particularly in the Dutch situation, it can be explained by large mobility between the tracks, in a sense that the better students, even when they originally entered a less requiring track, can quite easily change to a higher track (Jacobs and Tieben, 2009).

4.3 Ordered logit with unobserved heterogeneity

To make sure that our estimations are not biased by unobserved school characteristics, we estimate the ordered logit model with unobserved heterogeneity at school level. Similar as in the multinominal logit with unobserved heterogeneity, we assume in the ordered logit model with unobserved heterogeneity an additional intercept which captures the latent variable (for a methodological description, Rabe-Hesketh *et al.*, 2004). The results, estimated by the Stata Gllamm routine, are described in the third model of Table 5.

When we account for unobserved heterogeneity at school level, the outcomes differ only slightly from the original outcomes. In particular the influence of the motivation variables is further reduced as only 'teachers are doing their best at school' and 'time to start homework' remains significant. It is interesting to observe that if the students consider in the first class of secondary education that the teachers are doing their best, the number of later dropout students decreases. Having motivating teachers in the first class creates a benevolent school

[11] Robustness tests show that this result is also obtained when a class specific variable (cf. multilevel model) is added.

Table 5: Characteristics of dropout classes

	Without unobs. heterogeneity				Unobs. heter.	
	Model 1		Model 2		Model 3	
1. EXOGENOUS AND MOTIVATIONAL VARIABLES						
Gender (male = ref)	1.0383				1.0324	
	(0.2557)				(0.3771)	
Homeland mother (NE = ref)						
Morocco	0.9020				0.9654	
	(0.4067)				(0.7947)	
Surinam / Antilles	1.0731				1.0602	
	(0.5249)				(0.6289)	
Turkey	1.0953				1.1908	
	(0.3997)				(0.1464)	
Other	1.0242				1.0795	
	(0.7140)				(0.2779)	
Ability	1.0030				1.0027	
	-0.2208				(0.3092)	
Age	1.0418				1.0433	
	(0.2026)				(0.2288)	
"I like school"	0.9395	***	0.9457	***	0.9859	
	(0.0004)		(0.0009)		(0.4688)	
"Opinion about math	1.0089				0.9983	
teacher"	(0.7835)				(0.9627)	
"Teachers do their	0.9494	*	0.9405	***	0.9242	**
best at school"	(0.0635)		(0.0073)		(0.0102)	
Number of retentions	0.9853				1.0087	
	(0.7370)				(0.8577)	
"Teachers are pleased	0.9844				0.9768	
with my results"	(0.5301)				(0.3940)	
Truant	1.0634	*	1.0680	**	1.0027	
	(0.0651)		(0.0426)		(0.9448)	
"I pay attention during	0.9944				1.0314	
explantions"	(0.8193)				(0.2534)	
Start homework after	1.0528	***	1.0525	***	1.0362	***
	(0.0000)		(0.0000)		(0.0049)	
Education parents	0.9770				0.9760	
	(0.2177)				(0.2511)	
Work parents	1.0030				0.9959	
	(0.7706)				(0.7215)	
Attended parents' evening	1.0638				0.9988	
	(0.2073)				(0.9818)	
"A high degree is important"	1.0069				1.0012	
	(0.7443)				(0.9581)	
"Affiliation with homework"	0.9375	**	0.9435	**	0.9805	
	(0.0135)		(0.0172)		(0.4943)	
"Talking about school at home"	0.9956				0.9932	
	(0.8426)				(0.7790)	
Checking homework	1.0248				1.0016	
	(0.1850)				(0.9381)	

Table 5: Characteristics of dropout classes - continued

	Without unobs. heterogeneity				Unobs. hetero.	
	Model 1		Model 2			Model 3
2. CLASS CHARACTERISTICS						
Country side	0.8861	***	0.8824	***	0.9674	*
	(0.0000)		(0.0000)		(0.0908)	
Mean abilities in class	0.9491	***	0.9508	***	0.9374	***
	(0.0000)		(0.0000)		(0.0000)	
St deviation of abilities	0.9540	***	0.9548	***	0.8543	***
in class	(0.0000)		(0.0000)		(0.0000)	
Skewness abilities	1.0310		1.0356		1.1186	**
in class	(0.3597)		(0.2860)		(0.0158)	
Stdev age in class	2.9154	***	2.9562	***	6.5598	***
	(0.0000)		(0.0000)		(0.0000)	
Percentage of boys	1.9295	***	1.9428	***	4.6486	***
	(0.0000)		(0.0000)		(0.0000)	
Class size	1.0843	***	1.0825	***	1.1673	***
	(0.0000)		(0.0000)		(0.0000)	
Ethnicity in school	0.0969	***	0.0950	***	1.1307	
	(0.0000)		(0.0000)		(0.6117)	
In chosing a school,	0.7309	***	0.7315	***	0.4088	***
quality is important	(0.0000)		(0.0000)		(0.0000)	
School track	1.0000		1.0000		1.0001	***
	(0.2987)		(0.2175)		(0.0015)	
Cutpoint level 1	0.0326	***	0.0248	***	0.3397	**
	(0.0000)		(0.0000)		(0.0152)	
Cutpoint level 2	0.1464	***	0.1112	***	3.6595	***
	(0.0000)		(0.0000)		(0.0036)	
Cutpoint level 3	0.4422	***	0.3345	***	19.517	***
	(0.0067)		(0.0000)		(0.0000)	
Cutpoint level 4	1.4609		1.0935		111.4	***
	(0.2083)		(0.6907)		(0.0000)	
Cutpoint level 5	4.5479	***	3.3956	***	524.44	***
	(0.0000)		(0.0000)		(0.0000)	
Cutpoint level 6	10.9775	***	8.4743	***	1750.7	***
	(0.0000)		(0.0000)		(0.0000)	
Cutpoint level 7	35.5225	***	27.3666	***	8072.4	***
	(0.0000)		(0.0000)		(0.0000)	
Pseudo R^2	0.0629		0.0629		Var school	6.4045
X^2	2.9 e3		3.0 e3			(0.4466)
p-value	0.0000		0.0000			

p-values between brackets; ***, **, * denote significance at, respectively, 1, 5 and 10%

career. Concerning the class specific variables, all variables, expect for ethnicity, have a significant effect on the dropout decision. In the model with unobserved heterogeneity, both skewness and standard deviation of the abilities are significant but differ in sign. From the estimations we learn that some heterogeneity in abilities decreases the number of dropouts in the class (i.e., from the standard deviation in abilities), however, the smartest students should not dramatically outperform the other students (i.e., from the skewness). In other words, intelligent students motivate their class-mates, but highly gifted students have the opposite effect.

5 Policy conclusions

This paper examines the drivers of students who leave secondary education without a diploma at two levels: (1) what are the determinants behind the decision of the individual student to dropout?, and (2) why do some first year secondary education classes have multiple dropouts while others do not?

To analyse the first research question, we make a distinction between students who leave secondary education (a) with a diploma, (b) without a diploma, (c) in an apprenticeship program, or (d) those who were still at school by the end of the survey.

The results of the multinominal logit model reveal a significant influence of motivational factors on the individual decision to leave school without a diploma. This observation confirms previous findings in the literature (e.g., Dekkers and Driessen, 1997). Students' ideas and behavior, as well as influence from the environment (i.e., the parents and the school) are crucial. Students who like attending school, who have a favorable opinion about their teachers, who have a lower number of retentions and who pay attention during classes, have a lower probability of dropping out. Even controlled for unobserved heterogeneity in the sample, similar results are obtained.

Subsequently, it is interesting to examine what drives motivated students? Based on results for post-secondary education, we discuss the significant influence of the first year of secondary education. A simple tabulation shows that some first year classes do not have students who (later) dropout, while other classes have a large number (up to 7!) of dropouts. By using an ordered logit regression with and without controlling for unobserved heterogeneity, we can relate both individual variables (e.g., motivation, background, parents) and class variables (e.g., mean abilities, standard deviation abilities, class size, ethnicity) to the number of students dropping out of the class. It turns out that, controlled for individual characteristics of the students, parents and school, and for the individual motivation, first year class characteristics have an important impact on the number of students dropping out.

Similar to the review of Rumberger (2001), our findings suggest that policy makers can

influence the dropout decision at two interrelated levels.[12] A first intervention strategy could focus on the individual student's motivation. Triggering motivation of the individual student will reduce the number of dropouts. A second intervention alters the institutional environment. Particularly the first class of secondary education turns out to be extremely important. Intuitively, in the first year of education students with different background characteristics and with a different educational knowledge are put together. Our results suggest that it would be advisable to make smaller first year classes, with better (i.e., more motivating) teachers and within each class some outstanding students.

The outcomes of this study deliver additional research questions. Firstly, it would be interesting to focus on particular groups with higher dropout rates (e.g., based on the school track or ethnicity) and examine whether our results apply to these groups. Secondly, further research is needed to analyse the drivers of motivation and to see where in time motivation changes. We are particularly interested in motivation drivers which can be influenced by the school or policy makers. Thirdly, it would be interesting to complement the quantitative results by a qualitative research. A qualitative research could compare good and bad practice schools and test the validity of our empirical results. Finally, this research focusses on the characteristics of the first year of secondary education, but prevention strategies at that stage may already come late. As Rumberger (2001, p. 22) notes, "*dropout prevention strategies can and should begin early in a child's educational career*". Further research will definitely yield additional insights.

References

[1] Ainsworth-Darnell, J. and D. Downey (1998). Assessing the oppositional culture explanation for racial/ethnic differences in school performance. *American Sociological Review* 63, 536-553.

[2] Alexander, K., D. Enwisle and C. Horsey (1997). From first grade forward: early foundations of high school dropout. *Sociology of Education* 70 (2), 87-107.

[3] Astone, N. and S. McLanahan (1991). Family structure, parental practices and high school completion. *American Sociological Review* 56, 309-320.

[4] Astone, N. and S. McLanahan (1994). Family structure, residential mobility, and school dropout: A research note. *Demography* 31, 575-584.

[5] Bean, J. (1978). Dropouts and turnover: the synthesis of a causal model of student attrition. *Doctoral dissertation*, University of Iowa.

[12] Note, however, that this paper does not attempt to consider the causality of dropout influences. To do so, one would need an experimental setting or an instrumental variable approach, which are due to data limitations not available.

Table 6: Bivariate analysis

		With diploma %	Dropout %	Apprenticeship %	Stay in Sample %	Total
A. Characteristics students						
A1. Exogenous characteristics students						
Gender	Male	86.46	7.40	5.46	0.67	9,198
	Female	89.18	5.49	4.69	0.64	8,499
Ethnicity	the Netherlands	89.57	5.26	4.54	0.63	15,574
	Morocco	72.08	17.53	9.09	1.30	308
	Surinam / Antilles	76.42	13.35	9.38	0.85	352
	Turkey	72.38	18.41	8.44	0.77	391
	Other	75.14	14.74	9.36	0.76	1,058
Ability	Average	35.21		Minimum	6.00	
	Standard Deviation	11.34		Maximum	60.00	
A2. Ideas and behavior students						
A21. Push factors						
"At school, one	Totally disagree	74.67	16.00	9.33	0.01	150
can learn	Partly disagree	86.64	7.42	5.80	0.13	741
interesting things"	Partly agree	89.18	5.62	4.49	0.71	10,098
	Totally agree	88.15	6.01	5.25	0.59	5,806
"I like school"	Totally disagree	80.84	11.73	7.09	0.34	1,185
	Partly disagree	86.62	7.26	5.51	0.61	2,630
	Partly agree	89.36	5.40	4.46	0.78	6,408
	Totally agree	90.04	4.82	4.57	0.57	6,474
"Opinion about math	Average	2.74		Minimum	1.00	
teacher"	Standard Deviation	0.47		Maximum	4.00	
"Teachers do their	Totally disagree	69.06	17.96	12.98	0.00	362
best at school"	Partly disagree	79.72	11.61	8.36	0.30	2,308
	Partly agree	90.24	4.79	4.24	0.74	10,195
	Totally agree	91.56	4.00	3.78	0.66	3,624
Number of retentions	4 or more times	35.71	28.57	35.71	0.00	14
	3 times	44.44	33.33	22.22	0.00	27
	2 times	58.23	28.92	12.85	0.00	249
	1 time	78.59	13.32	7.91	0.18	2,718
	regular career	90.09	4.78	4.37	0.76	14,361
	1 year ahead	90.24	4.27	4.88	0.61	328
"Teachers are pleased	Totally disagree	75.17	14.57	9.77	0.50	604
with my results"	Partly disagree	82.04	10.27	7.27	0.42	2,133
	Partly agree	89.85	5.05	4.44	0.66	9,801
	Totally agree	91.05	4.38	3.88	0.69	4,045
"I pay attention during	(Almost) never	70.48	12.38	17.14	0.00	105
explanations"	Sometimes	84.15	9.46	5.96	0.42	2,600
	Often	89.24	5.50	4.65	0.61	9,665
	Always	90.27	4.55	4.31	0.88	4,222
A22. Pull factors						
Truant	Never	87.25	6.86	5.27	0.62	15,371
	Only few times	92.26	3.28	3.47	0.99	2,015
	Once every month	86.57	7.41	6.02	0.00	216
	Once every week	84.13	11.11	4.76	0.00	63
	About every day	58.33	25.00	16.67	0.00	12
Not concentrated during	Never	89.54	5.17	4.59	0.71	3,249
classes	Sometimes	89.48	5.42	4.45	0.64	10,710
	Often	85.07	8.13	6.21	0.59	2,190
	(Almost) always	78.33	11.59	9.87	0.21	466
Time spend on	Average	5.18		Minimum	1.00	
homework (in quarters)	Standard Deviation	2.10		Maximum	9.00	
Start homework after	Less 30 minutes	91.53	4.05	3.71	0.71	9,189
	1 hour	87.54	6.21	5.63	0.63	3,835
	2 hours	84.00	9.43	5.89	0.68	1,325
	3 hours	84.21	8.55	6.64	0.60	994
	4 hours	81.16	11.97	6.51	0.35	568
	5 hours	76.15	13.30	10.55	0.00	218
	More than 5 hours	71.63	17.70	10.67	0.00	356

Table 6: Bivariate analysis - continued

B. Characteristics environment						
B1. Exogenous characteristics parents						
Highest degree parents	Primary	80.19	12.76	6.56	0.50	1,615
	Lower secondary	83.09	9.77	6.56	0.58	5,730
	Higher secondary	91.77	4.01	3.54	0.68	6,756
	First step higher	91.79	3.01	4.45	0.74	2,559
	Second step higher	89.99	3.34	5.96	0.72	839
	Third step higher	86.87	3.03	8.08	2.02	198
Work of parents	Worker	87.56	7.45	4.62	0.37	4,822
	One-man business	86.33	7.13	6.02	0.51	1,361
	Self-employed	82.36	9.95	6.93	0.77	4,433
	Lower employee	91.46	4.41	3.43	0.70	1,722
	Middle employee	91.87	3.42	3.89	0.82	3,160
	Higher employee	91.50	2.89	4.78	0.83	2,177
B2. Interest in schooling and aspirations parents						
Attended parents' evening	None	89.80	6.09	3.70	0.41	4,250
	Some	92.41	3.29	3.58	0.71	13,448
"A high degree is important"	Not important	91.20	3.94	4.07	0.79	761
	Little important	92.57	3.62	2.93	0.88	2,045
	Important	85.49	8.12	5.79	0.60	10,889
	Very important	90.85	4.00	4.47	0.67	4,002
"Affiliation with homework"	Average	2.00		Minimum	1.00	
	Standard Deviation	0.60		Maximum	5.00	
"Talking about school at home"	Average	3.98		Minimum	1.00	
	Standard Deviation	0.70		Maximum	5.00	
Checking homework	Never	89.94	5.00	4.32	0.74	8,915
	Sometimes	88.15	6.17	5.12	0.55	5,054
	Frequently	86.06	7.67	5.79	0.48	1,865
	Almost always	81.54	10.61	7.44	0.41	726
B3. The neighborhood						
Urbanization	Very large city	78.56	11.84	8.55	1.04	1,824
	Large city	85.39	7.99	5.88	0.75	3,744
	Urban	86.63	6.91	5.80	0.66	3,793
	Rural	90.73	4.58	3.98	0.72	4,325
	Very rural	92.08	4.25	3.33	0.34	3,878

[6] Bowlby, J.W. and K. McMullen (2002). *At a Crossroads: First Results for the 18 to 20-Year-Old Cohort of the Youth in Transition Survey.* Statistics Canada.

[7] Bowles, S. (1972). Schooling and inequality from generation to generation. *Journal of Political Economy* 80, 219-251.

[8] Bridgeland, J., J. Dilulio and K. Morison (2006). *The silent epidemic - Perspectives of high school dropouts.* Civic Enterprises, Washington, p. 44.

[9] Bryk, A. and Y. Thum (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal* 26, 353-383.

[10] Cairns, R., B. Cairns and H. Necherman (1989). Early school dropout: Configurations and determinants. *Child Development* 60,1437-1452.

[11] Carbonaro, W.J. (1998). A little help from my friend's parents: Intergenerational closure and educational outcomes. *Sociology of Education* 71, 295-313.

[12] Coleman, J., E. Campbell, C. Hobson, J. McPartland, F. Mood, F. Weinfeld and R. York (1966). *Equality of educational opportunity.* Washington, D.C.: U.S. Government Printing Office.

[13] Cook, P. and J. Ludwig (1997). Weighing the burden of 'acting white': Are there race differences in attitudes toward school. *Journal of Policy Analysis and Management* 16, 256-278.

[14] De Graaf, P., N. de Graaf and G. Kraaykamp (2000). Parental cultural capital and educational attainment in the Netherlands. A refinement of the cultural capital perspective. *Sociology of Education* 73, 92-111.

[15] De Graaf, P. and H. Ganzeboom (1993). Family background and educational attainment in the Netherlands for the 1891-1960 birth cohorts. In Shavit, Y. and P. Blossfeld (Eds.). *Persistent inequality. Changing educational attainment in thirteen countries.* Boulder: Westview Press.

[16] Dekkers, H. and G. Driessen (1997). An evaluation of the educational priority policy in relation to early school leaving. *Studies in Educational Evaluation* 23, 209-230.

[17] DesJardin, S., D. Ahlburg and B. McCall (1999). An event history model of student departure. *Economics of Education Review* 18, 375-390.

[18] Dodde, N. and J. Leune (1995). *Het Nederlandse schoolsysteem.* Groningen: Wolters-Noordhoff.

[19] Ekstrom, R., M. Goertz, J. Pollack and D. Rock (1986). Who drops out of high school and why? Findings from a national study. *Teachers College Record* 87, 356-373.

[20] Epstein, J. (1990). School and Family Connections: Theory, Research, and Implications for Integrating Sociologies of Education and Family. *Marriage and Family Review* 15, 99-126.

[21] European Commission (2006). *Detailed Analysis of progress towards the Lisbon Objectives in education and training: Analysis of Benchmarks and indicators?* European Commission.

[22] Fernandez, R., R. Paulsen and M. Hirano-Nakanishi (1989). Dropping out among Hispanic youth. *Social Science Research* 18, 21-52.

[23] Gandara, P., K. Larson, H. Mehan and R. Rumberger (1998). *Capturing Latino Students in the Academic Pipeline*. University of Berkeley: Chicano/Latino Policy Project.

[24] Garnier, H., J. Stein and J. Jacobs (1997). The process of dropping out of high school: A 19-year perspective. *American Educational Research Journal* 34, 395-419.

[25] Gibson, M. (1997). Complicating the immigrant/involuntary minority typology. *Anthropology and Education Quarterly* 28, 431-454.

[26] Goldschmidt, P and J. Wang (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal* 36, 715-738.

[27] Greene, W. (2003). *Econometric Analysis*. Prentice Hall, New Jersey.

[28] Grisson, J. and L. Shepard (1989). Repeating and dropping out of school. In Sheppard, L. and M. Smith (Eds.), *Flunking Grades: Research and Policies on Retention*. New York: Falmer Press.

[29] Goldschmidt, P. and J. Wang (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal* 36, 715-738.

[30] Groot, W. and H. Maassen van den Brink (2007). The Health Effects of Education. *Economics of Education Review* 26, 186-200.

[31] Haan, P. and A. Uhlendorff (2006). Estimation of multinominal logit models with unobserved heterogeneity using maximum simulated likelihood. *DIW Berlin, Discussion Paper 573*, p. 19.

[32] Hanushek, E., J. Kain, J. Markman and S. Rivkin (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics* 18 (5), 527-544.

[33] Hanushek, E. and L. Wößmann (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal* 116 (510), 63-76.

[34] Hanushek, E. and L. Wößmann (2007). The role of education quality in economic growth. *The World Bank, Policy Research Working Paper* 4122.

[35] Haveman, R., B. Wolfe and J. Spaulding (1991). Childhood events and circumstances influencing high school completion. *Demography* 28, 133-157.

[36] Jacob, M. and N. Tieben (2009). Social selectivity of track mobility in secondary schools. A comparison of intra-secondary transitions in Germany and the Netherlands. Forthcoming in *European Societies*.

[37] Jansen, E. and M. Bruinsma (2007). *Students' first year experience with their learning environment: does it affect achievement*. In: Yurichenko, L. (ed.). Perspectives on Teaching and Teaching Issues. Nova Science Publishers.

[38] Jencks, C., M. Smith, M. Bane, D. Cohen, H. Gintis, B. Heyns and S. Michelson (1972). *Inequality: A reassessment of the effects of family and schooling in America*. New York: Basic Books.

[39] Jimerson, S. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology* 37, 243-272.

[40] Kalmijn, M. and G. Kraaykamp (2003). Dropout and downward mobility in the educational career: An event history analysis of ethnic schooling differences in the Netherlands. *Educational Research and Evaluation* 9, 265-287.

[41] Lee, V. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist* 35, 125-141.

[42] Lochner, L. and E. Moretti (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review* 94 (1), 155-189.

[43] Mainhard, T. (2009). *Time consistency in teacher-class relationships*. Doctoral Dissertation, University of Utrecht, p. 119.

[44] McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior, in Zarembka, P. (ed.). *Frontiers in Econometrics*, Academic Press, New York.

[45] McNeal, R. (1997). High school dropouts: A closer examination of school effects. *Social Science Quarterly* 78, 209-222.

[46] McNeal, R. (1999). Parental involvement as social capital: Differential effectiveness on science achievement, truancy, and dropping out. *Social Forces* 78, 117-144.

[47] Milligan, K., E. Moretti, P. Oreopoulos (2004). Does Education Improve Citizenship? Evidence from the United States and the United Kingdom. *Journal of Public Economics* 88 (9), 1667-1695.

[48] Pascarella, E. and P. Terenzini (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. *The Journal of Higher Education* 51 (1), 60-75.

[49] Pascarella, E. and P. Terenzini (1991). *How college affects students: findings and insights from twenty years of research.* Jossey–Bass, San Francisco, CA.

[50] Pong, S. and D. Ju (2000). The effects of change in family structure and income on dropping out of middle and high school. *Journal of Family Issues* 21, 147-169.

[51] Psacharopoulos, G. (2007). The costs of school failure - A feasibility study. *European Expert Network on Economics of Education.*

[52] Psacharopoulos, G. and R. Layard (1979). Human capital and earnings: British evidence and critique. *Review of Economic Studies* 46, 485-503.

[53] Rabe-Hesketh, S., A. Skrondal and A. Pickles (2004). Gllamm manual. *University of California, Berkeley Division of Biostatistics Working Paper Series*, paper 160, p.140.

[54] Roderick, M., J. Nagaoka, J. Bacon and J. Easton (2000). *Update: Ending Social Promotion.* Consortium on Chicago School Research, Chicago.

[55] Roemer, J. (1998). *Equality of Opportunity.* Harvard University Press, Cambridge.

[56] Rumberger, R. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research* 57, 101-121.

[57] Rumberger, R. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal* 32, 583-625.

[58] Rumberger, R. (2001). *Why students drop out of school and what can be done.* Conference proceedings "Dropouts in America: How Severe is the Problem? What Do We Know about Intervention and Prevention?". Harvard University, January 13, 2001.

[59] Rumberger, R., R. Ghatak, G. Poulos, P. Ritter and S. Dornbusch (1990). Family influences on dropout behavior in one California high school. *Sociology of Education* 63, 283-299.

[60] Rumberger, R. and S. Thomas (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education* 73, 39-67.

[61] Seltzer, M. (1994). Studying variation in Program success: A multilevel modeling approach. *Evaluation Review* 18, 342-361.

[62] Spady, W. (1970). Dropouts from higher education: an interdisciplinary review and synthesis. *Interchange* 1, 64-85.

[63] Steinberg L., P. Blinde and K. Chan (1984). Dropping out among language minority youth. *Review of Educational Research* 54, 113-132.

[64] Suichu, E. and J. Willms (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education* 69, 126-141.

[65] Swanson, C. and B. Schneider (1999). Students on the move: Residential and educational mobility in America's schools. *Sociology of Education* 72, 54-67.

[66] Tieben, N. and M. Wolbers (2008). Success and Failure in Secondary Education. Effects of family background on secondary school outcome. *Discussion paper for the RC28 spring meeting*, Florence, Italy.

[67] Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research* 45, 89-125.

[68] Train, K. (2003). *Discrete choice models using simulation.* Cambridge University Press, Cambridge, UK.

[69] Van Heusden Hale, S. (2000). *Comprehensive school reform: Research-based strategies to achieve high standards.* San Francisco: West Ed.

[70] Wilcox, P., S. Winn and F. Marylynn (2005). 'It was nothing to do with the university, it was just the people': the role of social support in the first-year experience of higher education. *Studies in Higher Education* 30 (6), 707-722.

TIER WORKING PAPER SERIES
TIER WP 10/01

ISBN 978-94-003-0001-9