Schools and Their Multiple Ways to Impact Students: A Structural Model of Skill Accumulation and Educational Choices

Annalisa Loviglio*

Job Market Paper

November 12, 2018

(click here for the latest version)

Abstract

This paper studies how the school environment affects students' cognitive skills and educational attainment. I estimate a dynamic structural model of cognitive skills accumulation and schooling decisions of students enrolled in lower secondary education, using rich administrative data for the universe of public schools in Barcelona. Its key feature is that it allows me to separately identify the different channels through which schools affect student outcomes. I find large variation across schools both in their effect on cognitive skills development, and in their effects on students' educational choices above and beyond their level of cognitive skills. School environment is particularly relevant for choices of students with disadvantaged family background. Moreover their probabilities of graduating or enrolling in upper secondary education if they attend a given middle school have limited correlation with their expected performance in that school. Results suggest that evaluating and comparing schools using only nation-wide assessments may not favor disadvantaged students, who particularly benefit from schools which increase educational attainment, not only test scores.

JEL codes: C35, I24, J24.

*Universitat Autònoma de Barcelona and Barcelona GSE. Correspondence: annalisa.loviglio@barcelonagse.eu. I am grateful to my supervisors Caterina Calsamiglia and Joan Llull for their guidance and continuous support, and to Peter Arcidiacono for useful conversations and advice. I would also like to thank Manuel Arellano, Ana Rute Cardoso, Arnaud Maurel, and participants at the IDEA Applied Working Group for their useful comments along different stages of the project. I am grateful to the staff at the Departament d'Ensenyament and IDESCAT, and in particular to Xavier Corominas and Miquel Delgado, for their help in processing the data. I acknowledge financial support from the La Caixa-Severo Ochoa Program for Centers of Excellence in R&D of the Spanish Ministry of Economy and Competitiveness.

1 Introduction

Higher educational attainment is associated with better labor market outcomes, and greater health and life satisfaction.[1] However, students' socio-economic background is often the main determinant of their educational prospects. How to provide inclusive and quality education which raises outcomes for all, particularly the most disadvantaged, is a long standing preoccupation for policy makers all around the world. Many countries have been implementing school accountability measures to monitor school quality, to take corrective actions, and in some cases to assign funding —the "No Child Left Behind" U.S. act of 2001 is a well-known example. In practice the measurement of school quality typically relies on the results of nation-wide assessments, with the underlying assumption that school ability of raising students' test scores is a sufficient measure of the school capability of improving individual outcomes in education.[2]

In this paper, I study how the middle school in which a student is enrolled affects performance, and the probabilities of graduation and of enrollment in academic upper secondary education. Results suggest that other metrics should be used together with test scores to effectively evaluate how schools increase educational attainment, especially among students with less favorable socioeconomic conditions. I find large variation across schools in their effect on cognitive skills development, but also in their effects on students' educational choices above and beyond their level of cognitive skills. Moreover, given that there is limited correlation between school effects in the different dimensions, being enrolled in a school with high value added on performance does not necessarily increase chances of pursuing further education. This is particularly relevant for subgroups of the population that are traditionally less likely to achieve high qualifications.

A large literature shows that life success depends on more than cognitive skills alone, and that interventions aimed at raising a broader set of skills have impressive returns in the long run, contributing to bridge the gaps due to family conditions.[3] These results emphasize that the debate around school quality should go beyond test scores alone.

[1]On average across OECD countries, the employment rate is 85% for tertiary-educated adults, 76% for adults with an upper secondary qualification, and less than 60% for those who have not completed upper secondary education. Moreover, 25-64 year-old adults with a tertiary degree earn 54% more than those with only upper secondary education, while those with below upper secondary education earn 22% less (OECD, 2018). Those with high literacy skills and a high level of education are 33 p.p. more likely to report being in good health than those with low literacy skills and a low level of education. 92% of tertiary-educated adults were satisfied with their life in 2015, compared to 83% with lower attainment (OECD, 2016).

[2]The "No Child Left Behind" act (replaced by the "Every Student Succeeds" act in 2015) requires public schools to administer a statewide standardized test annually; if school's results are repeatedly poor various steps are taken to improve the school. Since 1992, U.K. has been has published so-called "school league tables" summarizing the average GCSE results in state-funded secondary schools. Underperforming schools face various sanctions (Leckie and Goldstein, 2017).

[3]See for instance Cunha, Heckman, and Lochner (2006), Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010).

In fact, a child is left behind not only if she gets a low score in a standardized test, but also if she is not provided with the appropriate school environment to develop both her cognitive and non-cognitive skills, and to motivate her to pursue further studies. Secondary education is a crucial stage, because for the first time in their educational career students can choose whether they want to acquire further education and, in some cases, what they want to study. In fact, in most countries basic education is compulsory, but students are legally allowed to leave when they reach a given age, not upon completion of a given level. Moreover, in many European countries after completing lower secondary education students choose whether to enroll in the track which gives access to University. This decision is typically taken when they are 16 years old or even younger.[4] A student at risk of dropping out may be better off attending a school in which she feels comfortable and she is able to achieve a diploma, rather than another institution that would have potentially raised her final test score more, but from where she would have dropped out. Similarly, the choice of undertaking upper secondary or tertiary education may depend on previous performance, but also on student's motivation or family support.

In this paper, I estimate a dynamic model of cognitive skills accumulation and schooling decisions throughout lower secondary education of students enrolled in heterogeneous middle schools. At each time, cognitive skills growth depends on an unobserved ability, individual characteristics, and school environment (captured by classmates' characteristics and school effects). The student has imperfect information on her level of cognitive skills, but progressively learn about her true ability through various assessments. After updating her beliefs, she chooses whether to pursue further education. Her flow utility depends on her beliefs about cognitive skills, but also on the school environment and on individual characteristics. Importantly, before taking the decision, the student can be retained (i.e. required to repeat a level to stay in school). Retention may raise performance in the following period, but it increases the time needed to graduate and can change students' preferences.

Schools are heterogeneous and affect children in many dimensions. First, they differ in the way in which they contribute to the accumulation of cognitive skills for a given quality of peers. Second, they are different in their probability of retaining students with given cognitive skills and individual characteristics. Third, they influence students' educational choices in a different way. The main advantage of the structural approach is that it allows me to separately identify these different channels through the sequence of student decisions and test scores. Another advantage is that it allows me to quantify the relevance of informational frictions about own ability in explaining educational choices, which may be important to explain dropout decisions, especially among retained students

[4] For instance, such choice is typically made in the year in which the student turns 16 in Spain, 15 in France, and 14 in Italy.

since they received negative signals.

I estimate this model following an approach which builds on James (2011) and Arcidiacono, Aucejo, Maurel, and Ransom (2016). First, I estimate the grade equations, the variance of unobserved ability, and individual beliefs over time using an application of the Expectation-Maximization algorithm that makes the estimation computationally easier. Then, I estimate logit equations for the retention events. Finally, I estimate the parameters that govern the sequence of students' choices through maximum likelihood.

I employ administrative data on the universe of students attending lower secondary education in public schools in Barcelona (Spain) in the years 2009-2015. In this setting, nation-wide exams are administered at the end of primary education and at the end of lower secondary education, but the latter are measuring only a selected subsample, because a relevant amount of students dropout before taking the test. Moreover, given the compulsory education laws in Spain, all students spend at least some time in lower secondary education, and they are evaluated by their teachers at least once, even though these evaluations are not fully comparable across schools. Using the structure of the model, I can combine the signals provided by these different evaluations, even if they are not directly comparable across schools, and accounting for the self-selection of pupils into taking the standardized test.

Estimation results show that school environment is an important determinant of cognitive skills, both through peer effects (being with 1 standard deviation higher ability peers increases cognitive skills of more than 0.1 standard deviation) and through the school effect beyond peers (the interquantile range of school effect is 0.35 s.d.). Moreover school environment has a sizable direct effect on educational choices (being in a school at the 75 percentile of the distribution of school effects on the choice rather than at the 25 percentile increases flow utility from enrolling in high school as much as an increase of 0.6 s.d. in cognitive skills). Moreover results show that about half of the total variance of cognitive skills is due to unobserved ability. Evaluations are informative and students quickly acquire accurate beliefs.

I use the model to do three types of simulation exercises. In the first one, I evaluate the importance of school environment by simulating educational outcomes for different types of students in each of the schools in the sample. In particular, I focus on students with low educated and highly educated parents. For all types, the school environment is an important determinant of cognitive skills development. However for students with highly educated parents the school attended has relatively little importance on educational attainments, because they are extremely likely to graduate and enroll in high school regardless of the school environment. Conversely, for students with low educated parents the school attended can play a crucial role. In fact, the difference in graduation probability between the school at the 75 percentile and the school at the 25 percentile is more than 18 percentage points. I observe a similar gap for enrollment in upper secondary education. Moreover, the results of the simulation show that the correlation between predicted performance and probability of graduation is quite low: disadvantaged students have better chances to graduate in several schools with average predicted evaluations than in other schools with a potentially higher final test score.

In the second simulation exercise, I evaluate the impact of policies that raise school effects on cognitive skills or those on educational choices in schools below a given threshold. In all scenarios the overall graduation rate and enrollment in high school increase considerably, especially among students with low educated parents (up to 6 p.p. for graduation rate and 6.5 p.p. for enrollment in high school). However, on average, disadvantaged students enrolled in schools with high share of low parental background children may benefit more from interventions aimed at raising cognitive skills, while students enrolled in schools with low share of low parental background children may benefit more from intervention aimed at improving non-cognitive skills or tastes for education.

Finally, in the third simulation exercise, I study the consequences of retention. Although repeating a level raises cognitive skills, it has a net strong adverse effect on students' choice of pursuing further education. Retained students have somewhat lower beliefs on their ability than otherwise identical students promoted to next level. A counterfactual simulation without uncertainty about cognitive skills shows that most of the differences in their choices is due to changes in their utility, it is not due to the misperception of their ability.

This paper relates to several strands of the literature, particularly the literature on school quality, on human capital development, and on decision making. School accountability requires to develop reliable measures of school quality to compare among them schools (Allen and Burgess, 2013; Angrist, Hull, Pathak, and Walters, 2017; Kane and Staiger, 2002), or school types (e.g. charter *versus* traditional public schools, as in Dobbie and Fryer (2011) or Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011)). This is typically done with a value added approach, i.e. the estimation of the net effect of attending a given institution on a relevant outcome.[5] Test scores have been the most used outcome to capture "quality", under the assumptions that performances in school measures cognitive skills and are positively correlated with desirable outcomes in subsequent educational stages and in the labor market. This paper has a comparable "value added" approach, but it exploits a variety of outcomes, showing that school effects on

[5]Literature on school value added is closely related to the one on teachers value added, e.g. Chetty, Friedman, and Rockoff (2014).
Some of the most recent works investigate also the determinants of school value added. For instance Dobbie and Fryer (2013) and Angrist, Pathak, and Walters (2013) identifies practices such as increased instructional time, high-dosage tutoring, and high expectations which makes some charter schools particularly successful, and Fryer (2014) show that some of these best practices can be successfully exported to other school types.

performance and attainment are not aligned.

In fact, other lines of the literature have well established that cognitive skills alone do not explain educational choices and attainments which matter for all future life outcomes. On one hand, literature on human capital development has fully acknowledged that returns from non-cognitive skills are as high as the one from cognitive skills and the former may not be well captured using test scores (e.g. Heckman and Rubinstein (2001)). On the other hand, literature on decision making applied to educational choices shows that there is large variation among individuals with identical prior performances, due to a multitude of reasons, from differences in beliefs on the return of each choice, to different consumption values.[6] These works mainly focus on individual traits and preferences, and on their differences by gender and socio-economic background.[7] This paper contributes to incorporate their finding in the study of school quality. In fact it seems very plausible that school environment, similarly to the family environment, may substantially contribute to non-cogntive skills development, tastes formation, and provision of information on returns from education.[8] Having a structural model of cognitive skills development, retention, and choices allows me to first estimate the effect of school environment on cognitive skills, and then its direct effect on a given choice on top of the impact that goes through the effect on cognitive skills.

This is also what differentiates my work from other recent contributions (Angrist, Cohodes, Dynarski, Pathak, and Walters, 2016; Deming, Hastings, Kane, and Staiger, 2014) which exploit across schools differences in educational choices and attainment, such as high school graduation, college enrollment, college persistence. In fact they use those outcomes as measures of human capital alternative to test scores, for instance to show that the large gain in attending some charter schools found by previous works is not due

[6] Several papers study the choice of college major in US: Altonji, Arcidiacono, and Maurel (2015) surveys the literature. Avery, Hoxby, Jackson, Burek, Pope, and Raman (2006) and Hoxby and Avery (2012) study the role of financial constraints and information in applications to selective college of high-achieving students who are low income. Arcidiacono (2004) finds that individual preferences for particular majors in college or in the workplace is the main reason for ability sorting; Zafar (2013) and Wiswall and Zafar (2014) find that while expected earnings and perceived ability are a significant determinant of major choice, heterogeneous tastes are the dominant factor in the choice of major. On the other hand Wiswall and Zafar (2015) find that college students are substantially misinformed about population earnings and revise their earnings beliefs in in response to the information. Kinsler and Pavan (2015), Bordon and Fu (2015), Hastings, Neilson, and Zimmerman (2013) exploit Chilean data to study college and major choice.

[7] An interesting example is given by Belfield, Boneva, Rauh, and Shaw (2018). They collect survey data on students' motives to pursue upper secondary and tertiary education in UK. They find that beliefs about future consumption values play a more important role than beliefs about the monetary benefits and costs, and differences in the perceived consumption value across gender and socio-economic groups can account for a sizeable proportion of the gender and socio-economic gaps in students' intentions to pursue further education.

[8] Jackson (2018) shows that teacher value added on measures of non cognitive skills are important predictors of high school completion and college enrollment, even more than teacher value added on cognitive skills. Moreover the two values added are weakly correlated.

to a “teaching to the test” attitude, but to a true improvement of skills that matter in the long run. Their analyses cannot assess whether improvements in graduation rate or college enrollment are due to the improvement of cognitive skills as measured by standardized tests, or to other factors on top of that. The advantage of the structural model implemented in this paper is that it allows to disentangle school effect through cognitive skills and through other channels, and to assess how the two aspects interact.

I find that the school environment is particularly relevant for choices and attainments of students with low parental background or low prior cognitive skills. Dearden, Micklewright, and Vignoles (2011) question the use of a single measure of value added on performances to assess school effectiveness showing that schools can be differentially effective for children of differing prior ability levels. My paper shows that even under the simplifying assumption that performances of all students are affected similarly by a given school environment, schools may matter differently for educational attainments of students of differing background and ability.

Finally, my work contributes to the debate on the effectiveness of grade retention. Despite retention being a common practice in many countries, empirical literature provides mixed evidence of its effectiveness in improving student performances (Allen, Chen, Willson, and Hughes, 2009; Fruehwirth, Navarro, and Takahashi, 2016). My results suggest that it can improve test scores at the end of middle school (at the cost of longer time in education); however it has a negative effect of students' consumption value of schooling, therefore the net result is a large increase in dropout rate among retained students and lower probability of enrollment in high school.[9] Interestingly, the gap is larger for students of relatively higher ability, who would not be at risk of retention in schools that are more lenient.

The remainder of this paper is organized as follows. Section 2 provides background on the Spanish education system, describes the data, and discusses descriptive statistics. Section 3 describes the model and Section 4 details the estimation procedure. Section 5 presents the estimation results. Section 6 studies how the school environment affect educational outcomes of children with low or high parental background. Section 7 studies how raising school effects change average outcomes in the entire population and among students with low parental background. Section 8 concludes.

[9] Jacob and Lefgren (2009) and Cockx, Picchio, and Baert (2017) find that retention has adverse effect on probability of graduate from high school (in the USA and in the Flanders, Belgium respectively).

2 Data

I employ administrative data on the universe of students who began lower secondary education in a public middle school in Barcelona (Spain) in the years 2009 and 2010. I exploit various data sources to collect detailed information on enrollment, school progression, performances, and socio-demographic characteristics.
This section gives some background on the school system, describes the data sources, and discusses descriptive statistics.

2.1 Education system

In Spain basic education is divided into two stages: primary school (corresponding to ISCED level 10, primary education) and middle school (corresponding to ISCED level 24, general lower secondary education). Normally primary education takes 6 years, followed by 4 years of middle school. All students begin primary school in the year in which they turn 6 years old, ad they may repeat a grade at most once, thus they start middle school either in the year in which they turn 12 or one year after. In middle school, they may be retained at most twice, therefore they can graduate in the year in which they turn 16 years old or later on, depending on their previous attainments.

Students are legally required to stay in school until their sixteenth birthday, while they are allowed to leave school from the day after even if they did not complete lower secondary education. Given that retention is common, several students turn 16 well before the potential graduation date.

After successfully completing lower secondary education, students can enroll in high school for two more years (corresponding to ISCED level 34, general upper secondary education). They can also choose to attend vocational training, but only the former grants direct access to tertiary education after completion.

About 60% of students attend a public middle school. All public schools are largely homogeneous in infrastructure, curricula, funding per pupil, limit on class size, and teacher assignment. On the other hand, schools have large autonomy in deciding how to evaluate students' performances and whether to admit them to the next level.

Families have quite limited choices when it comes to select a middle school for their children. In fact, each primary school is affiliated with one or more middle schools: students from affiliated institutions have priority if the school is oversubscribed; to break ties other priority criteria such as the distance between school and home are used. The structure of the application process provide high incentive to put as top choice the school for which the student has the highest priority, because students who are not admitted in their first choice lose their priority for other schools. For instance, in 2009 92% of families apply to an affiliated middle school and 88% to the closest school.

2.2 Data sources and sample selection

The *Departament d'Ensenyament* (regional ministry of education in Catalonia) provided enrollment records for public schools in Barcelona, from primary school to high school, from the school year 2009/2010 to 2015/2016. In this paper I focus on the 44 schools which have available information for both the school year 2009/2010 and the school year 2010/2011.[10] For each year, data include school and class attended by the students, information on promotion or retention, final evaluations assigned by teachers at the end of the year. Moreover they contain information on time-invariant characteristics such as gender, nationality, and date of birth. They also allow me to identify children with special needs, which I drop from the sample.[11]

The *Consell d'Avaluació de Catalunya* (public agency in charge of evaluating the educational system) provided me with the results of standardized tests taken by all the students in the region attending 6$^{\text{th}}$ grade of primary school and 4$^{\text{th}}$ grade of middle school. Such tests are administered in the spring since 2008/2009 for primary school and since 2011/2012 for middle school. They assess students' competence in Maths, Catalan, and Spanish.[12] These exams are externally designed and graded. In this paper I refer to the test scores as *external evaluations*, in contrast with the final evaluations given by teachers in the school, which I call *internal evaluations*. The tests are administered in two consecutive days in the same premises in which students typically attend lectures. Normally every student is required to take all the tests, however children that are sick one or both days and do not show up at school are not evaluated.[13] In the analysis I use z-scores for both internal and external evaluations. I focus my analysis on students for whom I could retrieve the evaluations in primary school, i.e. 85% of the students who enroll in a public middle school in the period under analysis.[14]

Finally, information on the student's family background, more specifically on parental education, are collected from the Census (2002) and local register data (*Padró*). When the

[10]The IT infrastructure that supports the automatic collection of data has been progressively introduced since the school year 2009/2010. By year 2010/2011 most of the schools have already adopted it, while some are missing for 2009/2010.

[11]Special need children may have a personalized curriculum, and follow different retention rules. Therefore it would not be appropriate to include them in the estimation of my model. They are less than 3% of the total population of students who enrol for the first time in middle school.

[12]The tests are low stakes, because they do not have a direct impact on student evaluations or progress to the next grades but they are transmitted to the principal of the school, who forwards them to the teachers, families and students. More information can be found at: http://csda.gencat.cat/ca/arees_d_actuacio/avaluacions-consell (in Catalan)

[13]The school can also decide to exempt students with special educational needs and children that have lived in Spain for less than two years, but this is not relevant for the analysis given the sample that I am using.

[14]Evaluations can be missing for three reasons: 1. the student did not show up the day of the test; 2. the student did not attend primary school in Catalonia, she moved in the region only when she started middle school: 3. the student did take the test, but due to severe misspelling in the name or date of birth it was not possible to match the information with the enrollment data.

information can be retrieved from both sources, I impute the highest level of education, presumably the most up-to-date information. I allow for three level of education: Low (at most lower secondary education), Average (upper secondary education), High (tertiary education).

The online appendix contains a detailed description of the sample selection and the creations of the variables used in the analysis.

2.3 Descriptive statistics

The sample used for the analysis include 5000 students, who begin lower secondary education in September 2009 or in September 2010 in one of 44 public middle schools in Barcelona. About 17% of them do not graduate, i.e. they leave school before completing basic education; 9% dropout as soon as possible, while the other stay for one additional school year or more after reaching the legal age to dropout. 65.6% of the initial pool of students eventually enroll in high school (78.9% of those who graduated).

As shown in Table 1, there are wide differences across subgroups of the population. Children with low educated parents have much lower test scores when they start middle school, they are more likely to dropout at 16, and only 70% of them complete lower secondary education (the share is 95% among children with highly educated parents). Those who manage to graduate have on average lower test scores (-0.32 versus 0.59 for students with high parental background) and are less likely to pursue further studies. Only 42% of the initial pool of students with low educated parents enroll in high school, while 86% of students with highly educated parents do so.

There is a large gap also between students with immigrant background and natives: the former have lower performance and have lower probabilities of completing middle school or enrolling in high school.

Boys and girls have on average similar performance both at the beginning and at the end of middle schools; however there are significant differences in their attainments. Boys are more likely to dropout as soon as possible, 20% of them do not complete middle school (14% of girls), and only 60% enroll in high school (72% of girls).

Disadvantaged students are also more likely to have classmates with similar background; for instance, the average incoming test score of classmates of students with low educated parents is 0.5 standard deviation lower than the average. Conversely, boys and girls have similar peers.

The last four rows of Table 1 show descriptive statistics by retention status. Students are grouped into four categories: those who are never retained before leaving middle school, those who were retained in primary school (8%), those who are retained for the first time in middle school before reaching the last grade (15%), those who are retained for the first time in the last grade (4%). Students who were already behind before turning 16 years old are significantly more likely to be early dropout, especially if they have been retained during secondary education: 30% of them immediately leave schools, while only 3% of students with a regular progression dropout. Moreover less than half of them graduate and very few enroll in high school. Students who repeat the last grade are less likely to graduate and enroll in high school than the average but they have better odds than students retained at an early stage.

Table 2 describes the distribution of incoming students' characteristics and their outcomes at the school level; each column shows values of a given variable at various quantiles. It confirms that schools are quite different both in the types of students they teach and in the outcomes they produce. For instance, the school at the 75th percentile of average incoming test scores is more than 0.5 s.d. better than the school at the 25th percentile; the interquantile range of the share of students with low parental education is 19 p.p; in the school at the 90th percentile 81% of students enroll in high school, while only 44% do so in the school at the 10th percentile.

3 Model

3.1 Overview

I model cognitive skills development and educational decisions of students enrolled in middle school in Barcelona. Educational choices include staying in school after legal age to dropout is reached, and enrolling in further academic education. While in school, students may fail a level and have to repeat it: retention change their incentives to continue their education, especially because it prolongs the time required to achieve a diploma.

Cognitive skills accumulation depends on the level of cognitive skills in the previous period, individual and school characteristics, and an unknown (cognitive) ability. While in school, students receive evaluations that they use to infer their unobserved ability. There are two type of evaluations: 1. standardized grades, whose generating function is the same across schools; 2. internal grades, whose generating function may have school-specific components.

Retention is probabilistic and depends on student's cognitive skills, individual and school characteristics.

Students are assumed to be forward looking and choose actions which yield the highest utility. Their flow utility at each time depends on their beliefs on cognitive skills, on their individual characteristics, on the school environment, and their retention history.

3.2 Time line

The model mirrors the Spanish education system with some necessary simplifications.

- At the end of primary education, each student undertakes a nation-wide test. After completing primary education, they are assigned to a middle school and begin lower secondary education (time $t = 0$).
- Lower secondary education covers two levels (I and II). The normal length of a level is one time period, but students may be retained once, either during level I or during level II; in this case if they do not leave education they have to spend one more period in the same level.
- At time $t = 1$ students finish their first period in school; they receive internal grades and the communication of whether they have been promoted to level II. From next period education is not compulsory anymore, thus they have to decide whether to stay in school or dropout.
  - Retained students who continue in school repeat level I. At time $t = 2$ they receive new internal evaluations and they are surely promoted to level II.
  - Promoted students who continue in school access level II. At time $t = 2$ they receive internal evaluations and external evaluations from a nation-wide test, moreover they are informed of whether they successfully complete lower secondary education or they have been retained.
- At time $t = 2$ students who did not complete lower secondary education yet decide whether to leave school or to stay in level II. If they stay, at time $t = 3$ they receive internal and external grades; moreover they are told whether they graduate or not.
- Students who successfully complete lower secondary education (either at $t = 2$ or at $t = 3$) decide whether to enroll in high school to undertake further academic education.

3.3 Cognitive skills formation

The creation of cognitive skills is a cumulative process: the cognitive skills at a given point depend on the cognitive skills achieved in the previous level and on contemporaneous inputs. Students start secondary education with skills $C_{i,0}$, reach $C_{i,\mathrm{I}}$ at the end of the first level, and $C_{i,\mathrm{II}}$ at the end of the second level. When they repeat a level, the most recent knowledge replaces what learned in the previous time period. I denote $C_{i,\tau}(t)$ the cognitive skills in period τ at time t

The contemporaneous inputs are the unobserved ability h_i and a set of covariates $z_{it} = (x_i', st_{it}, p_{it}', s_{it}')'$, which include individual time-invariant characteristics x_i (e.g. gender, nationality, parents' education), a dummy st_{it} which takes value 1 if the level is repeated for the second time, a vector of peers characteristics p_{it} (e.g. average parental education of peers), and a vector of school dummies s_{it}. Peers at time t are the other students attending the same class; class composition may change overtime because sometime classes are shuffled at the beginning of a new school year, and because some students dropout.[15] On the other hand the middle school attended is constant from time $t = 1$.

$$C_{i,0} = z_0'\beta_0 + h_i \tag{1}$$

$$C_{\mathrm{I},it} = \alpha_{\mathrm{I}} C_{i,0} + z_{it}'\beta_{\mathrm{I}} + \mu_{\mathrm{I}} h_i, \qquad t \in \{1,2\} \tag{2}$$

$$C_{\mathrm{II},it} = \alpha_{\mathrm{II}} C_{i,\mathrm{I}} + z_{it}'\beta_{\mathrm{II}} + \mu_{\mathrm{II}} h_i, \qquad t \in \{2,3\}. \tag{3}$$

The cognitive ability h follows normal distribution $\mathcal{N}(0, \sigma)$; it is uncorrelated with z_{it} at any point in time. Students do not know h, while they know the cognitive skills production function.

3.4 Evaluations as signals

3.4.1 External evaluations

The nation-wide test score at time t in level τ is an unbiased measure of cognitive skills, i.e. it is an affine transformation plus an exogenous normally distributed error:

$$r_{\tau,it} = o_\tau + \lambda_\tau C_{i,\tau}(t) + \epsilon_{r_\tau,it}, \qquad \epsilon_{r_\tau,it} \sim \mathcal{N}(0, \rho_\tau^r). \tag{4}$$

The nation-wide test is administered only at the end of primary education and at the end of secondary education. Therefore, all students observe:

$$r_{0,i} = C_{i,0}(0) + \epsilon_{r_0}, \qquad \epsilon_{r_0,i} \sim \mathcal{N}(0, \rho_0^r), \tag{5}$$

and those who stay in school long enough also receive

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + \lambda_{\mathrm{II}} C_{i,\mathrm{II}}(t) + \epsilon_{r_{\mathrm{II}},it}, \qquad \epsilon_{r_{\mathrm{II}},it} \sim \mathcal{N}(0, \rho_{\mathrm{II}}^r), \tag{6}$$

[15] I define peers at the class level rather than a school level for two reasons. First, students in the same class are exposed to the same teachers and the same contents, spending all the school time together. Second, this allows p_t to vary both overtime and within school; given the limited number of cohorts I am analyzing this is a desirable feature.

with $t = 2$ or $t = 3$. Note that in period 0 the parameters (o_0, λ_0) have been normalized to (0,1).

3.4.2 Internal evaluations

At the end of each period in secondary education, students receive teachers' evaluations. Given that exams are designed and graded internally, teachers' biases or comparison with peers may affect the assigned score. Moreover schools may be more or less lenient, and administer more or less difficult tests. In other words, children with the same level of underlying cognitive skills may expect to receive different evaluations depending on their characteristics, peers, or school in which they are enrolled. Moreover, similarly to nation-wide test scores, there is an exogenous normally distributed error:

$$g_{\tau,it} = \nu_\tau + \mu_\tau C_{i,\tau}(t) + z_{it}'\gamma_\tau + \epsilon_{g_\tau,it}, \qquad \epsilon_{g_\tau,it} \sim \mathcal{N}(0, \rho_\tau^g). \tag{7}$$

Note that in principle all the contemporary observed determinants of cognitive skills can be a source of discrepancy between internal and external evaluations, while the unobserved ability h only affects evaluations through cognitive skills.

3.4.3 Identification of the grade equations

The scale factors in the grade equations, the shares α_τ, and the coefficients β_τ cannot be identified separately. Therefore, I will not be able to disentangle the contemporary effect of time invariant characteristics, but only their cumulative effect. Moreover a necessary assumption for identification is that school and teachers' policy for grading is constant across levels, i.e. $\gamma_{\mathrm{I}} = \gamma_{\mathrm{II}} = \gamma$.[16]
With some abuse of notation, I redefine evaluations as follow:

$$r_{0,i} = z_{i0}'\beta_0 + h_i + \epsilon_{r_0,i} \tag{8}$$

$$g_{\mathrm{I},it} = \nu_{\mathrm{I}} + z_{it}'(\beta_{\mathrm{I}} + \gamma) + \kappa_{\mathrm{I}} I_{0,i} + \mu_{\mathrm{I}} h_i + \epsilon_{g_{\mathrm{I}},it} \tag{9}$$

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + z_{it}'\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i} + \lambda_{\mathrm{II}} h_i + \epsilon_{r_{\mathrm{II}},it} \tag{10}$$

$$g_{\mathrm{II},it} = \nu_{\mathrm{II}} + \mu(z_{it}'\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i} + \lambda_{\mathrm{II}} h_i) + z_{it}'\gamma + \epsilon_{g_{\mathrm{II}},it}, \tag{11}$$

where $I_{\tau-1,i}$ is the portion of previous cognitive skills that comes from time-varying observed covariates.[17] The coefficients in β_τ capture the cumulative effects of time invariant regressors, and the innovation of time-varying regressors. Moreover, $\mu_{\mathrm{II}} = \mu\lambda_{\mathrm{II}}$.

[16] The latter assumption is necessary because external evaluations are observed only in level II.

[17] For instance, using the previous notation, for a student who did not repeat first level: $I_{i,\mathrm{I}} = p_{i\mathrm{I}}\beta_{p\mathrm{I}} + \alpha_{\mathrm{I}} s_{i0}\beta_{s0}$

3.4.4 Signals

I assume that students know the parameters that govern cognitive skills production function and grading, but they do not observe h_i, and therefore they do not know exactly $C_{i,\tau}(t)$ at any point in time. From the grades in school they infer signals on h_i and subsequently update their beliefs on their level of cognitive skills. More specifically,

$$s(r_{\tau,it}) = h_i + \frac{1}{\lambda_\tau}\epsilon_{r_\tau,it} \tag{12}$$

$$s(g_{\tau,it}) = h_i + \frac{1}{\mu_\tau}\epsilon_{g_\tau,it}. \tag{13}$$

All students observe $r_{0,i}$ and $g_{\mathrm{I},it}$, while the other signals they receive depend on their choices and on whether they are retained. After receiving one or more signals, students can compute the posterior distribution of their ability. When a new signal arrives, one can update the posterior distribution using the previous posterior as prior.[18]

For instance, suppose that a student of ability h is attending level II and receive both internal and external evaluations. Let s be the vector of signals, and μ, ω the prior mean and variance of h before observing s. Note that each signal has prior mean μ, and prior variance $\omega + \rho^e_{\mathrm{II}}$, $e \in \{r, g\}$. Then, from the point of view of the agent, (h, s') follow the multivariate normal distribution with mean values (μ, μ, μ) and variance covariance matrix $\begin{bmatrix} \omega & \omega & \omega \\ \omega & \omega + \rho^r_{\mathrm{II}} & \omega \\ \omega & \omega & \omega + \rho^g_{\mathrm{II}} \end{bmatrix} = \begin{bmatrix} \omega & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. Thus, the posterior distribution of h after receiving signals $s = \widehat{s}$ is simply the conditional distribution of h with normal distribution $\mathcal{N}(\overline{\mu}, \overline{\Sigma})$, where $\overline{\mu} = \mu + \Sigma_{12}\Sigma_{22}^{-1}(\widehat{s} - s)$ and $\overline{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

I use $\mathrm{E}_{i,t}(C_\tau)$ to denote the student's belief at time t about her cognitive skills in level τ. Moreover I denote $\psi_{it}(h)$ the posterior distribution after observing signals from time 0 to time t, and $\boldsymbol{\psi}_i(h)$ the final posterior distribution using all the available signals for i.

3.5 Retention and graduation

The events of retention and graduation are treated as probabilistic. In the first period everyone is at risk of retention and the probability depends on a set of characteristics $w_{i,1}$. I assume that the conditional probability takes a logit form:

$$\Pr(\mathrm{failI}_i = 1 | w_{i1}) = \frac{\exp(w'_{i1}\zeta_{\mathrm{I}})}{1 + \exp(w'_{i1}\zeta_{\mathrm{I}})}. \tag{14}$$

[18] See DeGroot (1970)

The set w_{i1} includes time invariant individual and peer characteristics, initial peers, middle school dummies, and prior beliefs $\mathrm{E}_{i,0}(C_I)$. This specification accounts for the fact that there is no deterministic rule in place to determine retention, in particular schools can choose to be more or less lenient. While I allow prior beliefs to enter the retention probability, I do not allow C_{iI} or equivalently h_i itself to enter the equation. This would make the model very difficult to treat, because students could learn about their ability through the realization of the event.[19] Moreover, this assumption appears sensible because the school personnel do not know either the true h_i when deciding about retention, but they can form a belief about it, exactly as the student does.

Individual are at risk of graduation only if they are in level II. For students who did not repeat period I the first failure is equivalent to a retention. Similarly to definition (14), I assume that the conditional probability takes a logit form:

$$\Pr(\mathrm{grad}_{it} = 1|w_{it}) = \frac{\exp(w'_{it}\zeta_{\mathrm{II}})}{1+\exp(w'_{it}\zeta_{\mathrm{II}})}, \qquad t \in \{2,3\}. \tag{15}$$

w_{it} includes individual and peer characteristics, middle school dummies, and prior beliefs $\mathrm{E}_{i,t-1}(C_{II})$. $Pr(\mathrm{grad}_{i2} = 1|w_{i2})$ can be set to 0 for individuals who are not in level II at time 2, i.e. who were retained in period I; for notational ease in next section I will use the expression $(\mathrm{grad}_{i2} = 0)$ also for those with $\mathrm{failI}_i = 1$.
Students are assumed to know the parameters and form expectations over their probability of graduation using (15).

3.6 Flow utilities

Students receive a flow payoff for each period they spend in school. This payoff depends on beliefs about the level of cognitive skills at beginning of the period, and on observable covariates y_{it}: history of retention ret_{it}, individual characteristics x_i, and school environment (peers' characteristics p_{it} and school dummies s_{it}). The specification includes interactions between beliefs on cognitive skills and groups of covariates. The coefficients capture all the motivational and non-cognitive factors which matter for the choice on top of the (perceived) level of cognitive skills.

The flow payoff of a period in lower secondary education for individual i attending

[19] In fact failI_i would be a signal with binary value and non-normal distribution. As a consequence the individual posterior distribution $\psi_{i1}(h)$ would not have a normal distribution. I follow Arcidiacono et al. (2016) in avoiding this complication.

level τ at time t is

$$
\begin{aligned}
U_{it}^{M} &= \phi_M \mathrm{E}_{i,t}(C_\tau) + \mathrm{ret}'_{M,it}\theta_{M,r} + x'_i\theta_{M,x} + p'_{it}\theta_{M,p} + s'_{it}\theta_{M,s} + && (16)\\
&+ \mathrm{E}_{i,t}(C_\tau)\Big[\mathrm{ret}'_{M,it}\kappa_{M,r} + \kappa_{M,x}(x'_i\theta_{M,x}) + \kappa_{M,p}(p'_i\theta_{M,p}) + \kappa_{M,s}(s'_{it}\theta_{M,s})\Big] + \varepsilon_{it} = \\
&= (\phi_M + \widetilde{y}'_{it}\kappa_M)\mathrm{E}_{i,t}(C_\tau) + y'_{it}\theta_M + \varepsilon_{it}, && (17)
\end{aligned}
$$

where $\widetilde{y}'_{it} = (\mathrm{ret}'_{M,it}, x'_i\theta_{M,x}, p'_i\theta_{M,p}, s'_{it}\theta_{M,s})$.

The vector $\mathrm{ret}'_{it} = (stI2_{it}, stII3_{it}, ftII3_{it})$ include three mutually exclusive dummies to capture all the possible combinations of time and repetitions. *stI2* takes value 1 if at time $t = 2$ if the student has to repeat the first level if she stays in school; *stII3* is 1 if the student has to repeat second level at time $t = 3$; *ftII3* is 1 for a student who repeated first level at time 2 and will undertake second level for the first time in period 3 if she stays in school. The specification in (16) allows cognitive skills to have different effects on the flow utility of students with differing retention history. This specification also allows school, peers, and individual characteristics to have different effects depending on the level of cognitive skills, while remaining parsimonious on the number of parameters to estimate.[20]

The flow utility for the choice of enrolling in high school has a similar formulation:

$$
\begin{aligned}
U_{it}^{A} &= \phi_A \mathrm{E}_{i,t}(C_{\mathrm{II}}) + \mathrm{ret}'_{A,it}\theta_{A,r} + x'_i\theta_{A,x} + p'_{it}\theta_{A,p} + s'_{it}\theta_{A,s} + && (18)\\
&+ \mathrm{E}_{i,t}(C_{\mathrm{II}})\Big[\mathrm{ret}'_{A,it}\kappa_{A,r} + \kappa_{A,x}(x'_i\theta_{A,x}) + \kappa_{A,p}(p'_i\theta_{A,p}) + \kappa_{A,s}(s'_{it}\theta_{A,s})\Big] + \varepsilon_{i,t} = \\
&= (\phi_A + \widetilde{q}'_{it}\kappa_A)\mathrm{E}_{i,t}(C_{\mathrm{II}}) + q'_{it}\theta_A + \varepsilon_{i,t}. && (19)
\end{aligned}
$$

$\mathrm{ret}'_{A,it} = (rII_{it}, rI_{it})$, with $rII = 1$ if the student repeated the second level, and $rI = 1$ if the student repeated the first level.

In each period the payoff of the outside option is normalized to 0 and the errors $\varepsilon_{i,t}$ are assumed to be logistic and i.i.d.

3.7 Choices and optimization

In each period students make a schooling decision taking in account their flow utility and expected future utility; individuals are assumed to be forward-looking and choose the sequence of actions which yield the highest expected value.

The one-period discount factor is δ. I use u_{it} to denote the utility at time t. Recall that students know all present and futures covariates while signals and shocks to preferences are random variables.

[20]For instance, with M schools in the sample, this specification requires the estimation of $M+1$ parameters to identify the school dummies coefficients and the interaction with cognitive skills. A specification with full interactions would require the estimation of $2M$ parameters.

At the end of lower secondary education. For those who graduated, the utility of pursuing further education is simply the flow utility in (18), with $t = 2$ if retention never took place or $t = 3$ if the student was retained in either first or second level. Therefore:

$$u_{it}|(\text{grad}_{i,t} = 1) = \max\left\{0\,,\ U^A_{it}\right\} = \max\left\{0\,,\ v^A_{it} + \varepsilon_{i,t}\right\}, \tag{20}$$

where v^A_{it} is the utility just before observing the realization of the random shock to preferences $\varepsilon_{i,t}$.

During lower secondary education. At $t = 2$ those who are still in education but did not graduate yet, repeat the choice of dropout, knowing that if they stay they will graduate with some probability and have the possibility to access upper secondary education.

$$\begin{aligned} u_{i2}|(\text{grad}_{i2} = 0) &= \max\left\{0\,,\ U^M_{i2} + \delta \Pr(\text{grad}_{i3} = 1)\mathrm{E}_{i,2}(u_{i3}|\text{grad}_{i3} = 1)\right\} = \\ &= \max\left\{0\,,\ v^M_{i2} + \varepsilon_{i2}\right\}. \end{aligned} \tag{21}$$

At $t = 1$ students make their first choice of dropout. They face different problems depending on the level that they will undertake if they stay in school. Those who are progressing regularly know that if they stay in school next period they may graduate with some probability or have to repeat level II. Conversely, those who are supposed to repeat level I anticipate that they will surely access level II in two periods if they stay, and then graduate with some probability:

$$\begin{aligned} u_{i1}|(\text{failI}_i = 0) &= \max\Big\{0\,,\ U^M_{i1}|(\text{failI}_i = 0) + \\ &+ \delta\Big(\Pr(\text{grad}_{i2} = 1)\mathrm{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 1)) + (1 - \Pr(\text{grad}_{i2} = 1))\mathrm{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 0)\Big)\Big\} \\ &= \max\left\{0\,,\ v^M_{i1}|(\text{failI}_i = 0) + \varepsilon_{i1}\right\}, \end{aligned} \tag{22}$$

$$\begin{aligned} u_{i1}|(\text{failI}_i = 1) &= \max\left\{0\,,\ U^M_{i1}|(\text{failI}_i = 1) + \delta\mathrm{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 0)\right\} = \\ &= \max\left\{0\,,\ v^M_{i1}|(\text{failI}_i = 1) + \varepsilon_{i1}\right\}. \end{aligned} \tag{23}$$

3.8 Identification

As common for this type of dynamic discrete choice models (e.g., see Rust (1987) and Arcidiacono et al. (2016)), identification of the flow utility parameters relies on the distributional assumptions imposed on the idiosyncratic shocks, the normalization of the outside option, and the discount factor δ, which is set equal to 0.95 throughout the paper.[21]

Under the assumptions on the parameters already discussed in Subsection 3.4.3, the identification of the grade equations relies on the assumption that educational choices only depend on students ability through their belief. In fact, evaluations at time $t \geq 2$ are only observed for individuals who chose to continue their education; to the extent that the choice depends on their ability, this raises a selection issue. The next paragraphs provide the intuition of why parameters of the grade equations can be consistently estimated.

Consider the following regression for external evaluations in level II and $t \in \{2,3\}$, based on equation (10) in Subsection 3.4.3. Let assume for now that posterior belief $\mathrm{E}_i(h)$ has been computed for each student.

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + z'_{it}\beta_{\mathrm{II}} + z'_{\mathrm{I},i}\widetilde{\kappa}_{\mathrm{II}} + \lambda_{\mathrm{II}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{r_{\mathrm{II}},it}, \tag{24}$$

where $\mathrm{E}_i(h)$ is the posterior ability for student i and can be expressed as a weighted sum of all the past ability signals. $z_{\mathrm{I},i}\widetilde{\kappa}_{\mathrm{II}} = \kappa_{\mathrm{II}} I_{\mathrm{I},i}$, where $z_{\mathrm{I},i}$ is the vector of time varying regressors from levels I and 0. Finally $\widetilde{\epsilon}_{r_{\mathrm{II}},it} = \lambda_{\mathrm{II}}(h_i - \mathrm{E}_i(h)) + \epsilon_{r_{\mathrm{II}},it}$. Under the assumption that educational choices depend only on posterior ability, the errors $\widetilde{\epsilon}_{r_{\mathrm{II}},it}$ is uncorrelated with regressors (i.e. $(h_i - \mathrm{E}_i(h))$ is white noise); therefore ordinary least square would consistently estimates the parameters $o_{\mathrm{II}}, \beta_{\mathrm{II}}, \lambda_{\mathrm{II}}$

Similarly, ordinary least square would consistently estimate the reduced form parameters of the following regression (based on equation 11:

$$g_{\mathrm{II},it} = \nu_{\mathrm{II}} + z'_{it}(\mu\beta_{\mathrm{II}} + \gamma) + z'_{\mathrm{I},i}\mu\widetilde{\kappa}_{\mathrm{II}} + \mu_{\mathrm{II}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{g_{\mathrm{II}},it}, \tag{25}$$

and using the previous estimates of β_{II} and λ_{II} one can retrieve estimates of μ and γ.

Having estimated γ, one could retrieve structural parameters β_I and μ_I from an application of ordinary least square to

$$g_{\mathrm{I},it} = \nu_{\mathrm{I}} + z'_{it}(\beta_{\mathrm{I}} + \gamma) + z'_{0,i}\widetilde{\kappa}_{\mathrm{I}} + \mu_{\mathrm{I}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{g_{\mathrm{I}},it}, \tag{26}$$

where $z_{0,i}\widetilde{\kappa}_{\mathrm{I}} = \kappa_{\mathrm{I}} I_{0,i}$ ($z_{0,i}$ are time varying regressors from level 0).

Finally, ordinary least square applied to

$$r_{0,i} = z'_{i0}\beta_0 + h_i + \widetilde{\epsilon}_{r_0,i} \tag{27}$$

allows to consistently estimate β_0 given that there is no selection at time 0. It is then possible to estimate $I_{0,i}$ and $I_{\mathrm{I},i}$ and retrieve the parameters κ_{I} and κ_{II}.

[21] I replicate the estimation using other values for δ in the interval $[0.9, 1)$ and results are virtually unchanged.

So far the identification of the parameters rested on the simplifying assumption that belief $\mathrm{E}_i(h)$ have been already computed. In fact, to perform the bayesian updating one should know the variance σ of the ability h and the variances of the errors in the grade equations $(\rho_0^r, \rho_{\mathrm{I}}^g, \rho_{\mathrm{II}}^g, \rho_{\mathrm{II}}^r)$. Those are identified from the history of signals, particularly the covarariance of evaluations within and overtime. In particular, σ is inferred from the covariance of the residuals of $g_{\mathrm{II},it}$ and $r_{\mathrm{II},it}$ on the observable regressors. The variance of each type of residuals is a linear function of σ and of the variance of the relevant error, thus the latter can be retrieved after estimating σ.[22]

4 Estimation

This section derives the likelihood of the model described in Section 3 and discusses its estimation.

4.1 Total individual likelihood

Let $d_i = (d_{it})_t$ (with $t \in \{1,2,3\}$) be the vector of choices of student i, $\mathrm{grad}_i = (\mathrm{grad}_{it})_t$ the vector of retention/graduation events, and $o_i = (o_{it})_t$ the vector of evaluations observed by i, where o_{it} contains one or two evaluations (in level II). The student takes $T_d \in \{1,2,3\}$ decisions, receives $T_{\mathrm{gr}} \leq T_d$ notification of retention/graduation, and observes signals T_d+1 times; more specifically she receives T_d internal evaluations and $T_r \geq 1$ external evaluations. For instance, consider a student who is retained in level I, stays one more period, and then dropouts; she takes two choices, $d_i = (1,0)$, she is notified retention once, and she observes $o_i = (r_{0,i}, g_{\mathrm{I},i1}, g_{\mathrm{I},i2})$.

Recall that $\boldsymbol{\phi}$ is the pdf of the random ability $h \sim \mathcal{N}(0,\sigma)$. Omitting for ease of notation the dependence on observable characteristics, the individual likelihood is

$$L_i = L(d_i, \mathrm{grad}_i, o_i) = L(d_{i1}, ..., d_{iT_d}, \mathrm{grad}_{i1}, ..., \mathrm{grad}_{iT_{\mathrm{gr}}}, g_{i1}, ..., g_{iT_d}, r_{i1}, ..., r_{iT_r}) \tag{28}$$

Moreover $L(d_i, \mathrm{grad}_i, o_i) = \int L(d_i, o_i|h)\boldsymbol{\phi}(h)dh$, therefore

$$\begin{aligned} L_i = &\int L(r_{i,0}|h)L(\mathrm{grad}_{i,1}|h, r_{i,0})L(g_{i,1}|h, r_{i,0})(L(d_{i,1}|h, r_{i,0}, g_{i,1})...L(d_{iT_d}|h, o_i, d_{i,1}, ..., d_{it_d-1})\,\boldsymbol{\phi}(h)dh = \\ &\Big(L(d_{i,1}|r_{i,0}, g_{i,1})...L(d_{i,T_d}|o_i, d_{i,1}, ..., d_{i,T_d-1})\Big) \times \Big(L(\mathrm{grad}_{i,1}|r_{i,0})...L(\mathrm{grad}_{i,T_{\mathrm{gr}}}|o_i, d_{i,1}, ..., d_{i,T_d-1})\Big)\times \\ &\times \int L(o_{i,T_d}|h, d_{i,1}, ..., r_{i,0}, ...)...L(r_{i,0}|h)\,\boldsymbol{\phi}(h)dh \end{aligned} \tag{29}$$

[22] More precisely, $\mathrm{Cov}(g_{\mathrm{II},it}, r_{\mathrm{II},it}|z_{it}, z_{\mathrm{I},i}) = \mu\lambda_{\mathrm{II}}^2\sigma$. For instance, the variance of the residual of $r_{\mathrm{II},it}$ is $\lambda_{\mathrm{II}}^2\sigma + \rho_{\mathrm{II}}^r$.

where the second equality follows from the fact that choices and retention/graduation probabilities depend on h only through students' beliefs, i.e. through the signals inferred from the evaluations. Thus the log-likelihood is separable in three parts (choices, retention probabilities, and evaluations), which can be estimated sequentially:

$$\log L_i = \log L_{i,d} + \log L_{i,\text{grad}} + \log L_{i,o} \tag{30}$$

Once coefficients of $\log L_{i,o}$ have been estimated, they can be used in the other components. In particular beliefs on cognitive skills can be computed for each student at any point in time and used as regressors. Then one has to estimate straightforward logit models for the probabilities, and a model of dynamic choices with logistic errors. More details are provided in subsections 4.3 and 4.4.

On the other hand maximizing the likelihood $\log L_{i,o}$ would be computationally costly because of the integration of h. Following James (2011) and Arcidiacono et al. (2016) I use an Expectation-Maximization (EM) algorithm to overcome this issue. I summarize the implemented approach in Subsection 4.2.

4.2 Cognitive skills

Let ζ be the vector of all the parameters that enter the grades equations (including variances of the idiosyncratic errors). Recall that $\boldsymbol{\phi}(h)$ is the density function of the unobserved ability, which follow normal distribution $\mathcal{N}(0, \sigma)$, and $\boldsymbol{\psi}_i(h) = \boldsymbol{\psi}(h|o_i; \zeta, \sigma)$ is the conditional density of h for individual i given her performances and the parameters.

For each individual i the likelihood $L_{i,o} = L(o_i; \zeta, \sigma)$ is the joint density function of the performances. To estimate the parameters (ζ, σ) one has to find

$$\arg\max_{\zeta,\sigma} \sum_i \log L(o_i; \zeta, \sigma) = \arg\max_{\zeta,\sigma} \sum_i \log \int L(o_i; \zeta, \sigma|h) \boldsymbol{\phi}(h)\, dh \tag{31}$$

The main point behind this application of the EM algorithm is that if $\widehat{\zeta}$ is a maximizer for (31), then it also solves

$$\arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma|h) \boldsymbol{\psi}_i(h)\, dh \tag{32}$$

Therefore for a given value of σ, $\widehat{\zeta}$ can be retrieved using (32) rather than (31). $\widehat{(\zeta, \sigma)}$ can be estimated using an iterative algorithm: at each iteration k, first (E-step) posterior distributions $\boldsymbol{\psi}_i^k(h)$ are estimated for all individuals using previous iteration estimates ζ^{k-1}. Then (M-step) estimates of pararameters ζ^k are computed as solution of (32).

Appendix (B) provides a more detailed theoretical motivation; next paragraphs describe the estimation procedure.

4.2.1 E-step

At step k, posterior distribution $\boldsymbol{\psi}_i^k(h)$ are computed for every students using all the observed evaluations, and the parameters $(\zeta^{k-1}, \sigma^{k-1})$ estimated in the previous iteration. Let $\mathrm{E}_i^k(h)$ be the individual posterior belief for h at iteration k and $\omega_i^k(h)$ the posterior variance. Moreover, at the end of E-step, I also update the estimate for the population variance; this new σ^k will be used at the beginning of next step $k+1$. The updating formula for σ^k is retrieved using the law of total variance:[23]

$$\sigma = \mathrm{E}\left(\omega_i(h) + \mathrm{E}_i(h)\mathrm{E}_i(h)'\right), \tag{33}$$

The sample equivalent at step k is computed as

$$\widehat{\sigma}^k = \frac{1}{N}\left(\omega_i^k(h) + \mathrm{E}_i^k(h)^2\right) \tag{34}$$

4.2.2 M-step

Given the individual posterior density functions $\boldsymbol{\psi}_i^k$ obtained in the E-step,

$$\arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh \tag{35}$$

can be solved to obtain an updated estimate ζ^k for the parameters in the evaluations equations. More specifically:

$$\sum_i \int \log L(o_i; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh = \tag{36}$$

$$= \sum_i \Big(\sum_t \int \log L(r_{it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \sum_t \int \log L(g_{it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh \Big) = \tag{37}$$

$$= \sum_i \int \log L(r_{0,i}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \sum_{it} \int \log L(g_{I,it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \tag{38}$$

$$+ \sum_{it} \int \log L(g_{\mathrm{II},it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \sum_{it} \int \log L(r_{\mathrm{II},it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh$$

[23] Let f be the vector of signals. Give that $\mathrm{E}(h) = 0$ and applying law of iterated expectations:

$$\mathrm{Var}(h) = \mathrm{E}(h \cdot h') - \mathrm{E}(h) \cdot \mathrm{E}(h') = \mathrm{E}(h \cdot h') = \mathrm{E}(\mathrm{E}(h \cdot h'|f)) = \mathrm{E}(\mathrm{Var}(h \cdot h'|f) + \mathrm{E}(h|f) \cdot \mathrm{E}(h|f))$$

where each sum is taken only on the relevant individuals and times. Given that the errors of the evaluations are normally distributed and the posterior distribution $\boldsymbol{\psi}_i^k(h)$ is known, the above expression can be derived as follow:[24]

$$
\begin{aligned}
-\sum_i logL_i \;=\; & \sum_i \frac{1}{2}\log(2\pi\rho_{r,0}) + \frac{1}{2\rho_{r,0}}\left(\omega_i^k + \left(r_{0,i} - (\mathrm{E}_i^k(h) + z_{i0}'\beta_0)\right)^2\right) + \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{g,\mathrm{I}}) + \frac{1}{2\rho_{g,\mathrm{I}}}\left(\omega_i^k + \left(g_{\mathrm{I},it} - \left(\nu_1 + \mu_{\mathrm{I}}\mathrm{E}_i^k(h) + z_{it}'(\beta_{\mathrm{I}} + \gamma) + \kappa_{\mathrm{I}} I_{0,i}\right)\right)^2\right) \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{r,\mathrm{II}}) + \frac{1}{2\rho_{r,\mathrm{II}}}\left(\omega_i^k + \left(r_{\mathrm{II},it} - \left(o_{\mathrm{II}} + \lambda_{\mathrm{II}}\mathrm{E}_i^k(h) + z_{it}'\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i}\right)\right)^2\right) \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{g,\mathrm{II}}) + \frac{1}{2\rho_{g,\mathrm{II}}}\left(\omega_i^k + \left(g_{\mathrm{II},it} - \left(\nu_{\mathrm{II}} + \mu\lambda_{\mathrm{II}}\mathrm{E}_i^k(h) + \mu(z_{it}'\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i}) + z_{it}'\gamma\right)\right)^2\right)
\end{aligned}
\tag{40}
$$

The total likelihood in (40) is the sum of four parts, one for each type of evaluations. Some students may contribute twice to the likelihood of an evaluation if they are retained, or they may not have some of them (if they dropout or do not take the final external evaluation).
If all the regressors were time invariant (i.e. if $I_{\tau,i}$ were not in the equations), the joint estimation of (40) would be completely equivalent to separately estimate the coefficients for r_0, then for g_{I}, and finally jointly estimates coefficients of r_{II} and g_{II}. Conversely the presence of time varying regressors makes all the four parts interdependent because past regressors have an indirect effect on evaluations in the following periods. Therefore a joint estimation is the most efficient. In practice, I found the following two-step MLE to be a good compromise between efficiency and speediness of the computations:

1. *Parameters for external evaluation at the end of primary school.*

    - Perform OLS regressions of $r_{0,i} - h_i$ over $z_{i,0}$. This provides us with updated estimates β_0^k, and allows the computation of $I_{i,0}^k = s_{i,0}\beta_{s,0}^k$, which is used in

[24] It is easy to see how to derive (40) from (38). For instance the contribution of the first nation-wide test is given by:

$$
\begin{aligned}
&\int \log L(r_{0,i};\zeta,\sigma^{k-1}|\eta)\psi_i^k(h)\,dh = \int \log\left(\frac{1}{\sqrt{2\pi\rho_0^r}}\exp\left(-\frac{(r_{0,i} - h - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0})^2}{2\rho_0^r}\right)\right)\psi_i^k(h)\,dh = \\
&= \int\left(-\frac{1}{2}\log(2\pi\rho_0^r) - \frac{1}{2\rho_0^r}(r_{0,i} - h - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0})^2\right)\psi_i^k(h)\,dh = \\
&= -\frac{1}{2}\log(2\pi\rho_0^r) - \frac{1}{2\rho_0^r}\mathrm{E}_i^k\left((r_{0,i} - h - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0})^2\right) = \\
&= -\frac{1}{2}\log(2\pi\rho_0^r) - \frac{1}{2\rho_0^r}\left(\mathrm{Var}_i^k\left(r_{0,i} - h - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0}\right) + \left(\mathrm{E}_i^k(r_{0,i} - h - x_i'\beta_{x,0} - p_i'\beta_{p,0})\right)^2\right) = \\
&= -\frac{1}{2}\log(2\pi\rho_0^r) - \frac{1}{2\rho_0^r}\left(\omega_i^k + \left(r_{0,i} - \mathrm{E}_i^k(h) - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0}\right)^2\right) \qquad (39)
\end{aligned}
$$

next step.

- Update variances $\rho^k_{r,0}$, using the sample equivalent of $\mathrm{E}_\epsilon(\mathrm{E}_h(\epsilon_{r_0,i}|r_{0,i}))$:

$$\mathrm{Var}(\epsilon_{r_0,i}) = \mathrm{E}(\epsilon^2_{r_0,i}) = \mathrm{E}(\mathrm{E}((r_{0,i} - h_i - z'_{i0}\beta_0|r_{0,i})^2)) = \tag{41}$$

$$= \mathrm{E}\left(\int (r_{0,i} - h - z'_{i0}\beta_0)^2 \, \psi_i(h)\, dh\right) = \tag{42}$$

$$= \mathrm{E}\left[\mathrm{Var}_i(h) + \left((r_{0,i} - \mathrm{E}_i(h) - z'_{i0}\beta_0)^2\right)\right] \tag{43}$$

which can be estimated from the sample as

$$\rho^k_{r,0} = \frac{\sum_i \left(\omega^k_i + (r_{0,i} - E^k_i(h) - z'_{i0}\beta^k_0)^2\right)}{N} \tag{44}$$

2. *Parameters for the other evaluations.* Maximize the joint likelihood of $g_{\mathrm{I}}, g_{\mathrm{II}}, r_{\mathrm{II}}$ using $I^k_{i,0}$ as a regressor.

4.3 Retention and graduation probabilities

Using the estimated parameters $\widehat{\zeta}$ and individual posterior distributions of ability $\widehat{\psi}_i$ at each point in time, we can compute beliefs on cognitive skills and use them to estimate retention and graduation probabilities, following the approach described in Section 3.5. I estimate two separate logit equations, one for retention in first period and one for graduation/retention in second period. The individual probability of graduation enters the student's maximization problem, while the probability of retention in first period does not, given that it happens before any decision has to be taken. However the estimation of the latter is necessary for simulation and conterfactual analysis.

4.4 Dynamic choices

We use the parameters of evaluations and probabilities to estimate the last piece of the model: the likelihood of the students' choices. It is important to recall that students use their *beliefs* on ability $\mathrm{E}_{i,t}(h)$ when they take a decision, not their true ability h_i; thus, when they compute their expected utility they anticipate that they will receive new signals and modify their beliefs. Therefore the computation of their expected utility for a given choice at time t requires to integrate over all the signals that they may receive from $t+1$ on. Their distribution is a multivariate normal, obtained through the usual bayesian updating. Let $\mathcal{N}(\widehat{h}_{it}, \omega_{it})$ be the (estimated) posterior distribution of h_i at t. Then $s(r_{\tau,it'})$, with $t' > t$, has posterior distribution $\mathcal{N}(\widehat{h}_{it}, \omega_{it} + \lambda_\tau^{-2}\rho_{r_\tau})$ and similarly $s(g_{\tau,it'})$ has posterior distribution $\mathcal{N}(\widehat{h}_{it}, \omega_{it} + \mu_\tau^{-2}\rho_{g_\tau})$; moreover, the posterior covariance of two signals is ω_{it}.

From now on, I will use $\widehat{\psi}_{it}(\mathbf{s})$ for the joint density function at t of a vector $\mathbf{s}$ of future signals. The updated belief $\widehat{h}_{it}$ is a linear combination of prior belief $\widehat{h}_{it-1}$, and contemporaneous signals $\mathbf{s}_t$; in other words there exists a vector of coefficients $\mathbf{c}_t$ such that $\widehat{h}_{it} = (\widehat{h}_{it-1}, \mathbf{s}'_t)\mathbf{c}_t$. The elements of $\mathbf{c}_t$ are functions of the elements of the covariance matrix and therefore are known to the agent. I will use this notation in the rest of this section to simplify the formulas.

I assumed that error terms ε_{it} are standard logistic, and uncorrelated with regressors and over time. It is well known that under these assumptions, the value of u_{it} just before observing the random shock to preferences ε_{it}, but knowing everything else, is

$$\mathrm{E}_\varepsilon(u_{it}|v_{it}^A) = \log\big(\exp(v_{it}^A) + 1\big) = \log\left(\exp(\phi_A \mathrm{E}_{i,t}(C_{\mathrm{II}}) + q'_{it}\theta_A + Hq'_{it}\kappa_A) + 1\right) \tag{45}$$

Recall that $\mathrm{E}_{i,t}(C_{\mathrm{II}}) = \mathrm{E}_{i,t}(h) + z'_{i,t}\beta_{\mathrm{II}} + k_{\mathrm{II}}I_{t-1}$. Given that in level II a student receives two signals, $\mathbf{s}_i = (s_{g,it}, s_{r,it})$, using the notation just introduced:

$$\mathrm{E}_{i,t}(h) = (\widehat{h}_{it-1}, \mathbf{s}'_t)\mathbf{c}_t = c_{0,t}\widehat{h}_{it-1} + c_{1,t}s_{r,t} + c_{2,t}s_{g,t} \tag{46}$$

Therefore, the ex-ante value in the previous period $t-1$ is

$$\begin{aligned} &E_{i,t-1}(u_{it}|\mathrm{grad}_{it} = 1) = \int \log\left(\exp(v_{i,A}) + 1\right) \cdot \widehat{\psi}_{it-1}(s_{g,t}, s_{r,t})\, d\mathbf{s}_t = \qquad (47) \\ &= \int \log\Big[\exp\left((\phi_A + \widetilde{q}'_{it}\kappa_A)\left(k_{\mathrm{II}}I_{t-1} + z'_{it}\beta_{\mathrm{II}}\right) + q'_{it}\theta_A\right) \cdot \exp\Big((\phi_A + \widetilde{q}'_{it}\kappa_A)(c_{0,t}\widehat{h}_{it-1})\Big) \cdot \\ &\cdot \exp\left((\phi_A + \widetilde{q}'_{it}\kappa_A)(c_{1,t}s_{r,t} + c_{2,t}s_{g,t})\right) + 1\Big] \cdot \widehat{\psi}_{it-1}(s_{g,t}, s_{r,t})\, d\mathbf{s}_t \end{aligned}$$

Moreover, the individual in period $t-1$ can compute $\widehat{\Pr}(\mathrm{grad}_{it} = 1)$ (the probability of graduating next period) using the estimated parameters for the probability of graduation and retention. This gives us a closed formula for v_{i2}^M:

$$v_{i2}^M(\widehat{h}_{i2}, z_{i2}) + \varepsilon_{i2} = U_{i2}^M + \delta\widehat{\Pr}(\mathrm{grad}_{i3} = 1)\int \log\big(\exp(v_{i3}^A) + 1\big) \cdot \widehat{\psi}_{i2}(s_{g,3}, s_{r,3})\, d\mathbf{s}_3 \tag{48}$$

Similarly, we are able to compute $\widehat{\Pr}(\mathrm{grad}_{i2} = 1)E_{i,1}(u_{i2}|grad_{i2} = 1)$. To conclude, we need to derive an expression for $\mathrm{E}_{i,1}(u_{i2}|\mathrm{grad}_{i2} = 0)$.

Again, thanks to the fact that errors are logistic, $\mathrm{E}_\varepsilon(u_{i2}|v_{i2}^M) = \log\big(\exp(v_{i2}^M) + 1\big)$ and $E_{i,1}(u_{i2}|\mathrm{grad}_{i2} = 0) = \int \log\big(\exp(v_{i2}^M) + 1\big) \cdot \widehat{\psi}_{i,1}(\mathbf{s}_1)\, d\mathbf{s}_1$. Finally, we can compute values for the first period:

$$v^M_{i,1}|(\text{failI}_i = 1) \; + \varepsilon_{i,1} = U^M_{i,1} + \delta E_{i,1}(u_{i2}|\text{grad}_{i2} = 0) \tag{49}$$

$$\begin{aligned} v^M_{i,1}|(\text{failI}_i = 0) \; + \varepsilon_{i,1} = U^M_{i,1} + \delta\Big(& \Pr(\text{grad}_{i2} = 1)\mathrm{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 1)) + \\ & + (1 - \Pr(\text{grad}_{i2} = 1))\mathrm{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 0)\Big) \end{aligned} \tag{50}$$

It is important to stress that from the point of view of a student in first period, $\widehat{h}_{i2} = (\widehat{h}_{i1}, \mathbf{s}'_2)\mathbf{c}_2$ is a random variable, and therefore it should be integrated out to compute the expectation. In v^M_{i2} it appears in the flow utility, in the continuation value from graduation, and in the probability of getting the diploma.

Finally, the likelihood of the individual choices can be easily retrieved computing probabilities with the usual formula for binary choices with logistically distributed preference shifters:

$$p_{it}(d_{it} = 1|d_{it-1} = 1) = \frac{\exp(v_{it})}{1 + \exp(v_{it})} \tag{51}$$

I maximize the total loglikelihood to estimate the parameters, following Rust (1987).[25]

4.5 Missing external evaluations

About 6% of the students in the sample who attain last grade did not undertake the external evaluation; this can happen if students are absent from school the day of the test. This possibility entails a small complication for my model: most students receive two signals in second level, but some only observe internal evaluations; they will therefore update their posterior beliefs differently. Moreover, when students (and the econometrician) compute expected utility they should take in account that with probability p they will observe two signals in last period, while with probability $1 - p$ they will observe only one signals. In practice I calibrate $\widehat{p}$ using the sample, and in the computation of expected utility I allow for the two different scenarios.

5 Results

In this section I present the results of the estimation. In subsections 5.1 to 5.3 I discuss the estimated parameters; in Subsection 5.4 I discuss the fit of the model. The main findings are the following:

[25]The integration over the signals is the most computationally costly part of the maximum likelihood estimation, it is performed using Gauss-Hermit quadrature.

- *Cognitive skills* About half of the total variance of cognitive skills is due to unobserved ability, however evaluations are informative and posterior individual variance shrink rapidly.[26] Parental education is the most important observed determinant of cognitive skills accumulation: having both parents with tertiary degree rather than primary education is associated with a 1 s.d. deviation improvement. School environment is quite relevant as well: being with higher ability peers increases evaluations (more than 0.1 s.d. for peers 1 s.d. above average), and there is large variation in the school effects (e.g. the difference between the school effect at the 75 percentile and the one at the 25 percentile is 0.35 s.d.).

- *Educational choices* The belief about cognitive skills is the most important determinant of choices, but also the school environment can have a large direct impact. For instance, being in a school at the 75 percentile of the distribution of the fixed effect rather than at the 25 affects the choice of staying in school as much as having 0.35 s.d. higher beliefs about cognitive skills; similarly, it increases flow utility from enrolling in high school as much as an increase by 0.6 s.d. of beliefs about cognitive skills. The school effect on the flow utility is non linear in cognitive skills: differences are larger for students at a lower level. Peer ability is associated with a decrease in the flow utility from staying in middle school (equivalent to 0.17 s.d. decrease in cognitive skills), perhaps due to ranking concerns.

- *Retention* School environment affects probability of retention and graduation on top of cognitive skills; for instance, being in a school at the 75 percentile of the distribution of the fixed effect rather than at the 25 affects probability of retention as much as having 0.35 s.d. higher beliefs about cognitive skills. When choosing whether to pursue further education, retained students exhibit a flatter flow utility in cognitive skills, therefore there is a large gap among retained and non-retained students with higher cognitive skills than with lower cognitive skills. Moreover, after completing middle school, the value of the flow utility of retained students relatively to their outside option is lower than the value of non-retained students at any level of cognitive skills; the gap is as large as a decrease of cognitive skills by 0.7 s.d.).

To simplify the notation, I use calligraphic letters in this and next sections to refer to the estimated coefficients of the vector of school dummies. $\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$ are school effects on cognitive skills in first and second level, $\mathcal{I}$ is the school effect on internal evaluation on top of cognitive skills (the "inflation effect"). $\mathcal{R}_{\mathrm{I}}$ and $\mathcal{G}_{\mathrm{II}}$ are school effects on retention in first level and graduation in second level. $\mathcal{T}_{\mathrm{I}}$ and $\mathcal{T}_{\mathrm{II}}$ are school effects on the choice of staying in middle school and the choice of enrolling in high school.

[26]The posterior individual variance is the updated variance of the student's belief after she receives one or more new signals.

5.1 Cognitive skills and evaluations

As reported in the first entry of Table 3 (Panel A), the estimated variance of the individual unknown ability is 0.278. As explained in Section 3.4.3, the total variance of cognitive skills in a given level does not have any direct interpretation; however, given the assumption that h is uncorrelated with the regressors, it is possible to decompose it in variance due to observable characteristics, and variance due to unknown ability. The remaining entries of Table 3 (Panel A) display the share of total variance in each level due to ability. In all levels about half of the variance in cognitive skills is due to the variance in ability; more specifically, about 49% before starting middle school and in the first level, and about 58% in the second level.

Panel B contains posterior variance in each period for every possible set of signals. Evaluations are quite informative, in fact the posterior variance at time $t = 1$ is just 0.06; at $t = 2$ it shrinks to 0.035 for retained students and 0.03 for students who are promoted and receive both internal and external evaluations; it is further reduced to about 0.02 for retained students who stay in middle school up to $t = 3$. Posterior variance is just slightly larger for students who did not undertake external evaluations in level II.

Table 4 presents estimates for the parameters governing internal and external evaluations. For the convenience of the reader, the table also reports separately the contributions of β and γ to internal evaluations in level I and II. Since the evaluations have been standardized, the coefficients of the individual characteristics can be interpreted as standard deviation changes of the relevant evaluation. Being a female is associated with a large premium for internal evaluations (0.29 in first level and 0.36 in second level), while boys and girls perform similarly in external evaluations (and thus there aren't relevant differences in their cognitive skills accumulation everything else equal). External evaluations are about 0.25 s.d. lower for non-Spanish students; the gap is negative also for internal evaluations although it is slightly smaller (about 0.18). Being younger when entering primary education is a disadvantage, although the gap is decreasing over time. Being born at the beginning of January rather than at the end of December is associated with an increase of about 0.25 s.d. in the test score before starting middle school and with a increase of about 0.12 s.d. at the end of it. Being retained in primary school, i.e. starting middle school with a one-year delay is associated with a large disadvantage in performances (up to -0.66 in external evaluations at the end of middle school).

Results show that parental background is a fundamental determinant of cognitive skills. Having better educated parents is associated with large improvements in performances, with quite similar effect for internal and external evaluations; for instance, having a mother with tertiary rather than primary education increases performances of about 0.5 s.d., having a highly educated father increases performances of about 0.4 s.d.. This means that a student with highly educated parents is expected to receive evaluations almost 1 s.d. higher than a classmate with identical characteristics and unobserved ability, but whose parents only have primary education.

Repeating a level for a second time has a positive effect on cognitive skills, especially in second level, In fact final external evaluations are about 0.4 s.d. larger for retained students everything else equal (the effect is 0.17 in first level).

The empirical analyses use a polynomial of degree 3 with orthonormal components for each peer regressor. To facilitate the interpretation of the results, Table 4 shows the effect on evaluations associated with having peers 1 s.d. above the mean value rather than at the mean. Peers' average evaluations at the end of primary school, an *ex ante* measure of average peer quality, is the peers' characteristic which affects cognitive skills development the most; 1 s.d. increase of this measure is associated with more than 0.1 s.d. increase in external evaluations at the end of middle school. On the other hand, the positive effect is completely offset in internal evaluations. This is aligned with the finding in Calsamiglia and Loviglio (2018) that teachers' evaluations are deflated in the presence of better quality peers, i.e. that teachers are somewhat "grading on a curve". Average parental education among peers has a positive effect on cognitive skills in the first level (about 0.1 s.d. deviation), while the effect is positive but quite small in second level. The other peer characteristics appear to have limited effects.

To summarize the estimated school effects, Table 4 displays the difference between the school dummy coefficients at the 75th percentile and at the 25th percentile, and the difference between school dummy coefficients at the 80th and 20th percentile. There is a sizable variation across schools in their effects; the interquantile range for the effect on evaluations through cognitive skills is 0.35 s.d. in the first level and almost 0.2 in second level. The interquantile range for the inflation γ_S of internal evaluations is sizeable as well (0.21). However the school inflation effect $\mathcal{I}$ and the school effects on cognitive skills ($\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$) have large negative correlation (about -0.7). In other words, schools which improve the cognitive skills the most tend to have stricter grading policy.[27] Therefore the correlation of the ranking of schools based on school effect on external evaluations and total school effect on internal evaluations is quite low, although positive.

5.2 Retention and graduation

Estimates of the parameters governing the two logit models for retention in first level and retention/graduation in second level are reported in Table 5.[28] Not surprisingly, higher

[27]This is coherent with the the evidence in Calsamiglia and Loviglio (2018) that schools in which average external evaluations are higher exhibit stricter grading policy.

[28]The table has the same structure of Table 4: for peer regressors it shows the effect on the average student in the sample of increasing the variable of one standard deviation; for school effects it reports

cognitive skills decreases the probability of retention in both periods. Being female and having better educated parents decrease the probability of retention in both periods above and beyond the cognitive skills level.[29]

There is sizable variation in school effects on retention, and they are correlated across levels, i.e. in schools that have a stricter retention policy in the first level it is also more difficult to graduate in the second level. For retention in the first level, the effect of being in a school at the 25 percentile rather than at the 75 percentile is the same as increasing cognitive skills by 0.34 s.d.. For graduation, being in a school at the 75 percentile rather than at the 25 percentile has the same effect of increasing skills of 0.5 s.d..[30]

School effects on retention are negatively correlated with school effects on internal evaluations (-0.6 in first level and -0.7 in second level), in other words schools which have more generous grading policy are also less likely to retain students.

5.3 Flow utilities

Table 6 presents the estimates of the flow utility parameters. Beliefs about cognitive skills have a large positive effect on both the choice of continuing middle school and the choice of enrolling in high school.

Peers' average incoming test score has a negative effect on the choice of staying in middle school (beyond their positive contribution to cognitive skills development). This may be related to ranking concerns: for a given level of cognitive skills, students at the margin of dropout may have a further reason to leave school if they dislike being among the worst in the class.[31]

School effects exhibit sizable differences both for the choice of staying in middle school and the choice of enrolling in high school. For the choice of staying in middle school, being in a school at the 75 percentile of school effects rather than at the 25 percentile impacts the flow utility as much as improving cognitive skills by 0.45 s.d. Given that the coefficient of the interaction between cognitive skills and school effects is almost 0, this difference is about constant for all level of cognitive skills. For the choice of enrolling in high school, the difference between being in a school at the 75 percentile rather than at the 25 percentile is equivalent to an increase of cognitive skills by 0.67 s.d. for someone with average level of cognitive skills. However the interaction coefficient is negative, therefore

differences between percentiles of the distribution.

[29] Students behavior in class may affect the decision of retention. Moreover teachers may communicate with parents during the school years, inform them that the child is at risk of retention, and to some extent take in account their preferences. This might explain the results.

[30] These figures are computed using the total variance of cognitive skills reported in Table 3.

[31] Findings in Pop-Eleches and Urquiola (2013) corroborate this interpretation. They find that being with better peers have positive effects on achievements, but children who make it into more selective schools realize they are relatively weaker and feel marginalized.

the gap is wider for children with below average cognitive skills and smaller for above average students. Figure 1 helps clarifying this point. The figure plots the school effects at each level of cognitive skills: schools with a higher coefficient for the dummy have a larger intercept and a smaller slope. The point in which the lines intersect is about 2 standard deviation above the average value of cognitive skills at the end of second level: students with such a high level of cognitive skills are extremely likely to enroll in high school anyway, therefore it is sensible to focus on the part of the graph which lies to the left of the intersection. There, at any given level of cognitive skills, students have a lower payoff from the choice of enrolling in high school if they attended a school with low effect. This gap increases when cognitive skills decreases. In other words, schools affect the flow utility of students in the lower tail of the distribution of cognitive skills the most.

All dummies for the retention events have negative coefficients. Interestingly, their interactions with cognitive skills also have negative coefficients, thus the total coefficient of cognitive skills for retained students is smaller than for students who are progressing regularly. In other words, being retained decreases the weight that individuals give to cognitive skills when making their choices. Figure 2 illustrates how the retention status affects the effect of cognitive skills on flow utility of enrollment in high school. At any level of skills retained students have a lower payoff, but the gap is is wider at higher levels of cognitive skills: an increase in skills improves less the utility of retained than of non-retained students.[32] Being retained in first or in second level have quite similar effects. As discussed in previous Subsection 5.2, retention probability is decreasing in cognitive skills, however many students with cognitive skills below average, but not particularly low, face a sizable probability of being retained at some points, especially if they are male; retention may particularly discourage those guys from pursuing further education.

5.4 Fit of the model

To assess the fit of the model, I simulate choices and outcomes of each individual in the sample, using the structural parameters estimates presented in the above sections. More specifically, I create 100 copies of each individual at time 0 (i.e. of her time invariant characteristics and primary school attended). For each of them I draw ability and shocks to evaluations, preferences, and retention events, using the estimated distributions; I can then compute their outcomes, cognitive skills beliefs, and choices.

[32] It is important to recall that the dummies for retention also capture any difference in the value of the outside option for retained and regular students. If retained students have better outside options (for instance because being older it is easier for them to find a job) then the negative gap between the values for regular and retained students would be captured by the coefficient of the dummies. Thus, the fact that the lines for for retained students lie below the one for regular students do not mean that they like high school less in absolute terms, but that they value it relatively less compared with the outside options.

Table 7 reports empirical frequencies ("data" columns) and frequencies predicted using the model ("model" columns) for the following events: choice of staying in school at time 1, graduation, enrollment in high school, retention in first level, retention in second level. Frequencies are computed over the entire sample at time 1. Table A-14 in the appendix reports similar statistics computed only on the subsample of the initial population who reached the relevant stage for the event to take place (e.g. enrollment on high school on the subgroup of students who completed middle school). The first line of each table contains frequencies for the overall sample, while the following rows contain the same type of information by relevant subgroups (e.g. parental background, quality of peers...). The predicted choice of staying in school, graduation rate, and choice of enrolling in high school are very close to the empirical one, both in the overall sample and by subgroups. The retention rate in first level is also similar to the empirical one, while the retention rate in second level is somehow higher (about 1.7 p.p. more).

Next, I investigated how evaluations and events simulated by the model replicate the patterns observed in the data. For instance, Figure 3 plots the share of students who chose to stay in school at $t = 1$ by quantile of their test score at $t = 0$; Figure 4 plots their enrollment in high school, again by quantile of the initial test score. The model predictions mimic the empirical outcomes quite well. Other evaluations and choices exhibit similar patterns.

6 School environment and parental background

In this section I study how the school environment affects outcomes of children with different parental background. In particular I focus on students with low educated parents (both attained at most lower secondary education) and on students with highly educated parents (both have tertiary education). As discussed in Section 5.1, having more educated parents is associated with a larger growth of cognitive skills, even if the level of unobserved ability is the same. Table 8 shows that on average students with low and high parental background experience also a different school environment.

In particular, panel A shows that students with low educated parents have peers with less educated parents, who are more likely to be non-natives, and to have experienced retention in the past. Looking at the distribution of mean peers' evaluation at the end of primary school, the expected value for a student with low educated parents is at the 22 percentile, while it is at the 64 percentile for children of highly educated parents. Similarly, the share of immigrant classmates is at the 70 percentile rather than at the 40.

Panel B displays average school effects for the two parental backgrounds; the typical school for a student with low educated parents has a somewhat more generous grading policy, a smaller propensity to retain in first level, larger direct effect on the choice of staying in school, and contribute slightly less to cognitive skills accumulation in first level. Average values of school effects for second level are close enough for the two types. Figure 5 confirms that there is some positive correlation among school effect on choice of staying and share of students with low educated parents enrolled in the school, while the share of students with low educated parents is uncorrelated with school effect on choice of high school, and has small negative correlation with school effect on cognitive skills. Overall differences in average school effects appear less relevant than those for peers, possibly with the exception of school effect on dropout. Importantly, almost all schools have both students with low educated and high educated parents.[33]

Results in Subsection 5.1 show that both peer quality and school effects on top of observable peer characteristics matter for cognitive skills accumulation. Given the linearity of the model of cognitive skills accumulation, effects due to school environment are the same for all students in a given class. Therefore if on average school environment affect differently students with low or highly educated parents, this is due to their different allocation across schools.

On the other hand, subsections 5.2 and 5.3 show that schools affect students' attainments and choices beyond their contribution to cognitive skills. School environments that are not particularly effective in boosting cognitive skills may have a positive effect on choices of pursuing further studies: in fact, school effects on cognitive skills and choices are negatively correlated, and for a given level of cognitive skills students are more likely to dropout if they have higher performing peers. Importantly, school environment may affect differently classmates who differ in their individual characteristics or ability. This is due to the non linearity of the probability functions: the effect on the probability of a change in one of the regressor depends on the initial level of the underlying utility.[34] Moreover, for the choice functions the flow utilities as well are non linear, because they include in-

[33]Overall 17% of students in the sample have both parents with at most lower secondary education, and 16% have both parents with tertiary education. All schools have at least 2% of students with low educated parents (more than 6% in 41 schools) and all but 2 have at least some students with high educated parents (more than 6% in 33 schools).

[34]As explained in Section 4.4, given the assumptions that shocks to preferences follow a logistic distribution, the probability of undertaking choice d_t at time t is $\Lambda(v_t(z)) = \frac{\exp(v_t(z))}{1+\exp(v_t(z))}$, where z is the vector of individual and school variables and v is the expected utility (i.e. the flow utility at time t before observing the shock to preferences plus the expected utility from the future). The marginal effect of a change in one of the regressor z_k is

$$\frac{\partial \Lambda(v_t(z))}{\partial z_k} = \frac{1}{1+\exp(v_t(z))} \frac{\exp(v_t(z))}{1+\exp(v_t(z))} \times \frac{\partial v_t(z)}{\partial z_k}$$

$\frac{\exp(v_t(z))}{(1+\exp(v_t(z)))^2}$ has its maximum in $v_t(z) = 0$ (which corresponds to a probability of 0.5), while it goes to 0 for $v_t(z) \to \pm\infty$. So the marginal effect is greatest when the probability is near 0.5, and smallest when it is near 0 or near 1. Thus for instance if $v_t(z)$ is very large not only it is very likely that the student undertakes the choice, but the probability would also be almost unaffected by small changes in the variables.

teractions between school environment and cognitive skills and between retention history and cognitive skills (and schools differ in their propensity to retain students). Therefore, the school environment may matter differently for educational choices of students with low or highly educated parents even if they are attending the same class.

I use the model and the estimated parameters to study performance, choices, and attainments of students with low and high parental background in their typical school environment or in counterfactual environments (Subsection 6.1). Then, I simulate outcomes of a given student in each school of the sample to quantify the variance in performances and attainments due to the school environment (Subsection 6.2). Finally, I explore differences in educational patterns of retained and non-retained students who have identical cognitive skills before retention (Subsection 6.3).

More specifically, for a given set of individual and school variables, I create 10000 fictitious students, and I draw shocks to evaluations, preferences, and retention events. I can then compute the educational pattern of each agent and estimate expected outcomes of an individual with a given set of individual characteristics and school environment. To make results fully comparable, I keep constant all individual characteristics beside parental education; in particular the reference characteristics for the results discussed in this paper are: male, Spanish, born in the middle of the year, began middle school at 12 years old; the unobserved ability is set to its average value of 0.[35] For brevity, I will call *type L* the students with such characteristics and low educated parents, and *type H* the students with such characteristics and high educated parents.

6.1 Peer and school effects

Table 9 summarizes outcomes of type L and type H when they are in their typical school environment, and under various counterfactual scenarios in which elements of the school environment of type H are assigned to type L and vice-versa. In the baseline simulation (columns (1)) each type attends a school with the average peers and school characteristics of his type. The first column of the table reports also results for a student with average parental background for comparison. Not surprisingly, type L receives below average internal and external evaluations throughout his studies, while type H's performances are above average. For instance, they enter middle school with external evaluations of -0.35 and 0.50 respectively, and those who graduate score on average -0.21 and 0.69 respectively. Type L students have 59% chances of being retained in first period, and almost 15% of

[35] I also assign the average primary school effect in the sample, and the average cohort effect. I replicated the analysis discussed in this paper with different set of characteristics, particularly for female and immigrant students. Overall results exhibit similar patterns, with the expected differences due to the alternative characteristics; for instance, female are less likely to be retained, and they choose to pursue further education more, immigrants are more likely to dropout.

dropping out immediately; more than 30% of them do not complete middle school and only 25% overall enroll in high school (about 36% of those who complete middle school). Conversely, only 4% of type H students is retained in first period, less than 1% dropout immediately, and only 1% do not graduate; about 94% of students enroll in high school (95% of those who graduate).

The remaining columns of Table 9 show results of conterfactual simulation in which a given type is enrolled in a school with typical school effects of the other type (columns (2)), or peers of the other type (columns (3)), or both (columns (4)). The comparison of the outcomes in the counterfactual and in the baseline specification allows us to gain a deeper understanding of the role plaid by the school environment on students' outcomes.

When type L attends a school with the average school effects of type H, he acquires a slightly larger level of cognitive skills – the final evaluation is about 0.05 s.d. higher. However he is slightly more likely to be retained, and he choose to dropout more both at time $t = 1$ and at time $t = 2$. Therefore he is about 5 p.p. less likely to graduate (graduation rate is 0.64 rather than 0.69); probability of enrolling in high school conditional on graduation has a small increase of 1.5 p.p. Type H would have a small drop in cognitive skills if he attended the typical school of the low type, but his attainments would be virtually unaffected. Changes in the grade inflation coefficient $\mathcal{I}$ and in the cognitive skills coefficients $\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$ largely compensate each other, therefore for both types internal evaluations are almost unchanged.

A change in peers composition has more dramatic effects, as shown in columns (3) of the table. For type L, the improvement in parental background and test score of peers implies an increase of cognitive skills of more than 0.3 s.d. at time $t = 1$. A slightly lower but still sizable improvement takes place also at the end of middle school. Conversely internal evaluations are similar in first level and somewhat lower in second level because the negative effect due to "grading on the curve" more than compensate the improvement in cognitive skills. New peers decrease the probability of retention, which is reduced by half. The boost in human capital more than compensate the negative effects that having better performing peers has on the choice of remaining in education; in fact dropout rate plummets from about 15% to about 7% at time $t = 1$ and from 23% to 15% at time $T = 1$, and the graduation rate raises from about 69% to about 84%. Moreover enrollment in high school conditional on graduation increases from 36% to 52%.

Type H experiences a symmetric drop in performances, and an increase of retention probability at $t = 1$ from 4% to 13%. On the other hand, dopout rate at $t = 1$ changes from 0.7% to 2.2%, graduation rate from 99% to 97%, enrollment in high school from 95% to 90%: the underlying flow utilities for this type are so high that variations in them lead to relatively small changes in probabilities.

Estimated outcomes in columns (4), where a type is given the overall average environment of the other type, are close enough to those in column (3); changes in cognitive skills are even more pronounced because peers and school effects go in the same direction, while changes in attainments are slightly smaller.

Summarizing, type L would benefit from attending the typical school in which type H is enrolled mainly because "better peers" would increase his cognitive skills, and this would have a large positive impact on his choices and attainments. The improvement in cognitive skills would not raise internal evaluations due to the counterbalancing effect of "grading on the curve". Conversely, the average school environment beyond peers is not very different for the two types, and it would not dramatically change type L outcomes; if anything, on average the school which he is already attending is better suited to increase his motivation to acquire further education on top of its level of cognitive skills. However, results in Section 5 show that there is large variance across schools in their school effects beyond peers, therefore the same type of student may experience a quite different educational path depending upon the specific institution in which he is enrolled. I will study the relevance of school environment school by school in next Subsection 6.2.

Results in this section and parameters estimated in Section 5 suggest that the event of retention can be important for subsequent choices and attainments; Subsection 6.3 will provide further insights.

6.2 Variation of outcomes across schools

To quantify the importance of the school environment on attainments and choice, I use the model to predict for each school the outcomes that a given type of student would have if enrolled there. Table 10 presents results for type L and type H. As usual, for each type I created 10000 fictitious students, and I drew shocks to evaluations, preferences, and retention events; then I simulate their outcomes in each of the 44 schools in the sample. I use the average peer characteristics in the school for the peer regressors; results are therefore representative of the outcomes that a student of a given type would have enrolling in one of the 44 schools in the sample. For comparison, I also focus on peer and school effects separately: I perform a simulation in which peer characteristics are set at their average values in the sample (Table A-18) and a simulation in which school effects are set at their average in the sample (Table A-19). Note that in all these simulations type L and type H are exposed to exactly the same environment in a given school: this allows me to quantify how relevant is the school environment for each of the type.

The left half of Table 10 contains results for the student with low parental background, the right half contains results for the student with highly educated parents. The first column of each part contains median outcomes across school. Patterns are quite similar to those discussed in previous Section 6.1; for type L median outcomes are slightly more favourable than the mean values in column (1) of previous Table 9, because each school has the same weight in the computation of the median value across schools, although they have different share of students with low educated parents.[36]

The remaining columns of Table 10 give us a sense of how different those figures are across schools. Each column contains the difference between the expected outcome at the (100-p) and at the p percentile, with $p \in 20, 25, 30$. Here I will comment the results using the interquantile range (i.e. $p = 25$). Given the linear model for cognitive skills, differences in evaluations across schools are quite similar for the two types.[37] Differences are important: during the first level, students acquire cognitive skills 0.55 s.d. larger in the school at the 75 percentile than in the school at the 25 percentile (045 s.d. larger in the second level).

For type L, variation in graduation prospects across schools is quite large: the interquantile range is almost 19 p.p. This reflects sizable differences in dropout rate across schools. Moreover the large differences in retention rate may have important second order effects on graduation, because retained students are more likely to leave school at time 1 and face the choice of dropout again at time 2. Interquantile ranges are definitely lower for type H: for instance it is 2 p.p. for graduation. In fact, variation of the underlying utilities across schools is similar for the two types, but given that their levels are much higher for type H, a unitary change in flow utility affects way less the probability of type H than the one of type L.

School environment also have a large effect on type L's probability of enrolling in high school (the interquantile range is about 18 p.p.). The effect is smaller but not negligible for type H (the interquantile range being 7 p.p.).

Tables A-18 in the appendix shows that for the low type there would be large differences across schools even if peers were evenly distributed. Moreover difference in peers are associated as well with large variation in outcomes across school (Table A-19). Peers and school effects do not move outcomes in the same direction: for a given outcome the correlation between average values simulated with constant peers or with constants school effects is typically low or negative.[38]

The rankings of schools based on the probability of dropout, graduation, or enrollment in high school are almost the same for the two types. However, for most of those events

[36]More specifically, for type L the median retention rate in first level is 45% , the median probability of dropping out immediately is more than 11% , that of not completing middle school is more than 25% , while chances of enrolling in high school are less that 33% (about 45% conditional on graduation). Conversely, the median retention rate of type H is less than 7% , 1% for dropout at $t = 1$, and less than 2% for not completing middle school; moreover the median rate of enrollment in high school is about 91% (93% conditional on graduation). In the median school the final evaluations among those who graduate are -0.15 and 0.59 respectively.

[37]They are not exactly the same due to the different selection of students in second level

[38]It is small and positive (0.14) for cognitive skills, while it is negative for outcome in middle school (-0.3 for the probability of graduation) and negligible (-0.05) for the enrollment in high school.

differences in probabilities across schools are so low for type H that moving from a top ranked school to one in the bottom tail would not dramatically change his prospects. Conversely, for type L the school environment can determine large changes in his expected attainments. Moreover the results of the simulation show that there is little correlation between schools' capability of increasing performances and their capability of leading everyone to graduation, or motivating students to choose further academic education. For instance, Figure 6 plots for each school the probability of graduation of type L (y-axis) and his predicted external evaluations at the end of middle school (x-axis). Although the two measures have a small positive correlation (0.22), there is clearly a lot of dispersion: many schools with average predicted evaluations have higher rates of graduation than schools with much larger predicted evaluations.

6.3 Retention and its consequences

Results in Section 5 show that repeating a level has a positive effect on cognitive skills accumulation, but it can also have a direct negative effect on educational choices, increasing probability of dropout and decreasing the probability of enrolling in high school, for a given level of cognitive skills. Moreover, as discussed in Section 5.2, both peers and school effects can impact the probability of retention. I now quantify the effect of repeating first or second level, comparing outcomes of students with identical characteristics, but different retention history.

Table 11 presents result for the type L student, using the average school environment among students with low educated parents as in column (1) of Table 9.[39] Retention at time 1 increases cognitive skills at the end of first level by about 0.2 s.d. Retention at time 2 for students enrolled in second level increases cognitive skills at the end of second level by about 0.4 s.d. Despite these positive effects on cognitive skills, retained students are by far less likely to complete middle school and enroll in high school. In particular, the ex-ante probability of graduation for a student retained at time $t = 1$ is about 56%, while for an identical student regularly promoted to next level it is more than 86%; moreover the ex-ante probability of enrolling in high school for the retained student is less than 13%, while it is about 42% for an identical student who did not experience retention. Graduation rate at time t =2 for students who were not previously retained is 86%; students who fail at time 2 have more than 16% probability of dropout, therefore, despite having very high probability of graduating if they stay in school next period (almost 93%), they have a lower graduation rate of 78%. Moreover their probability of enrolling in high school is only 30% despite the large increase in cognitive skills.

[39] To create the table I used the output of the baseline simulation discussed in previous Subsection 6.1, averaging outcomes by retention status.

There are two potential main drivers of this discrepancy. First, retention has a direct negative effect on flow utility for choices of pursuing further education. Both the fact that retained students like less to be in school or the fact that they have better outside options (or a combination of the two) are compatible with the estimated parameters.

Second, as shown in Table 11 which compare real and *perceived* cognitive skills, retained students have lower beliefs than identical students who were promoted: at time $t = 1$ the true level of cognitive skills for type L is -0.57, but the average perceived value for a retained student is lower (-0.61), while it is higher for a promoted student (-0.53). Similarly at time $t = 2$, the true value for a student in second level is -0.5, but on average student who graduate have a perceived value of -0.45 while those who are retained on average believe that it is -0.57. Thus retained students are more likely to underestimate their true ability, and given that choices are based on beliefs, they would be more likely to dropout even in the absence of any direct negative effect of retention on their utility. On the other hand, students who complete middle school at time $t = 3$ have both actual and perceived level of cognitive skills higher than those of identical students who graduated at time $t = 2$, thus the large gaps in enrollment rate are surely due to differences in preferences.

To understand how important the uncertainty on own ability is in explaining the different choices of retained and regular students, I replicate analysis in Table 11 under the conterfactual scenario in which students' ability h is perfectly known rather than unobserved. Results in Table A-15 show that removing uncertainty about ability would reduce dropout rate at the end of the period in which the student is retained by 1-2 p.p., increasing by a similar amount the probability of completing middle school. It would also slightly decrease the probability of enrolling in high school conditional on graduation (from about 50% to about 48%) for students who complete middle school at time $t = 2$, given that on average they slightly overestimate their cognitive skills in the baseline simulation.[40] Overall, the comparison of outcomes with and without uncertainty on ability shows that retained students are somewhat penalized by the randomness of the signals, but most of the differences in choices and attainments is due to changes in preferences.

Finally, Table A-17 presents a drill down by retention status analogous to the one in Table 11, but using type H student. Outcomes for type H follows a similar pattern than those for type L. In fact, although he still has much better prospects than type L even when he repeats, retention in first level is associated with a 10 p.p. drop in his probability of graduating, while retention in second level with a 5 p.p. drop. Moreover his probability of enrolling in high school after graduation is largely impacted, because it shrinks from

[40] For comparison, Table A-16 replicates the analysis in Table 9 without uncertainty. Average outcomes are close enough with and without uncertainty.

95% to 76% (retention in first level) or 79% (retention in second level). It is, however, worth to recall that retention is a concern for a relatively small fraction of type H students (about 4% of them are retained in first level and only 1% in second level).

7 Counterfactual improvements of school effects

I now use the estimated parameters to simulate students' choices and outcomes under counterfactual scenarios in which some of the school effects are increased. In particular, I study the effects of improving school effects on cognitive skill, and the effect of raising school effects on either the choice of staying in school or the choice of enrolling in upper secondary education, or both. As in previous sections, I first study changes on aggregate outcomes, and then focus on students with low parental background.

These counterfactual simulations can be regarded as government interventions targeting schools where a given school effect is below some threshold. Such interventions keep peer quality constant and act on school resources and personnel beyond peers. For instance, improving school effects on cognitive skills can be thought as hiring more qualified teachers or implementing remedial classes to strengthen the knowledge of the core subjects tested in the final evaluations. On the other hand, school effects on choices on top of cognitive skills may be improved providing students at risk of dropout with additional counseling to motivate them to remain in school, or mentoring students close to graduation on the broader opportunities they would gain if they acquire a high school diploma. These exercises abstract from the costs that the interventions would entail, but allow to quantify outcomes of interventions that would involve a similar number of schools with the goal of homogenizing them in one or more dimensions.

7.1 Simulations using the entire sample

Table 12 summarizes outcomes under five counterfactual simulation; in all of them the targeted school effects are raised at the 75 percentile value if they are lower. The column "baseline" contains average outcomes under the benchmark model. In column $(\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}})$ school effects on cognitive skills are modified. Column $(\mathcal{T}_{\mathrm{I}})$ simulates an intervention on school effects on choice of staying in school, while in column $(\mathcal{T}_{\mathrm{II}})$ school effects on choice of high school are improved; then in $(\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}})$ both interventions are simultaneously implemented. Finally, in column $(\mathcal{R}_{\mathrm{I}})$ school effects on retention at time $t = 1$ are adjusted. The table shows graduation rate, rate of enrollment in high school, average level of cognitive skils, and other outcomes under the baseline and the counterfactual scenario.[41]

[41]Note that in each counterfactual up to a quarter of schools (those with school effects above the 75 percentile) are not interested by the intervention, therefore their outcomes do not change. Computing

The intervention on $\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$ raises average cognitive skills of more than 0.2 s.d. at time $t = 1$, while it improves them by about 0.1 s.d. among graduated students. The improvement in average cognitive skills has a sizable indirect effect on graduation rate, which improves from 83% to 87%, thanks to a decrease in dropout rate (from 8% to 6% at time $t = 1$) and an improvement in graduation probabilities. Moreover the enrollment in high school among graduates raises by more than 2 p.p.: the larger pool of graduate students and the growth of the enrollment probability among them raise the unconditional probability of enrolling in high school by 5 p.p.

By construction, interventions on "tastes" for education do not change average cognitive skills of the overall sample, and they can only decrease the average among graduated students if they avoid dropout of the less able students. In fact, a potential concern of implementing interventions aimed at improving other aspects than cognitive skills is that they might keep in school or prompt to enroll in additional education students who do not have the necessary competences for it. Results in Table 12 suggest that this should not be a concern: interventions which raises $\mathcal{T}_{\mathrm{I}}$ or $\mathcal{T}_{\mathrm{II}}$ only cause a negligible decrease in the average cognitive skills after graduation, although they are associated with sizable increase in the graduation rate. More specifically, improving $\mathcal{T}_{\mathrm{I}}$ raising the lower values at the 75 percentile reduces dropout rate by 2 p.p. at time $t = 1$ (from 8.4% to 6.2%) and more than 3 p.p. at time $t = 2$ (from 17.6% to 20.9%). Graduation probabilities among this larger pool stay constant at time $t = 1$ and only slightly decrease at time $t = 2$, therefore the overall graduation rate improves of 2 p.p. Overall enrollment in high school is unaffected. Improving school effects on the choice of enrolling in upper secondary education ($\mathcal{T}_{\mathrm{II}}$) reduce dropout during middle school (of about 1 p.p. both at time 1 and 2), thus increasing graduation rate, because it increases the future expected utility from graduation. Moreover it raises enrollment in high school of 4 p.p.

Interestingly, raising simultaneously $\mathcal{T}_{\mathrm{I}}$ and $\mathcal{T}_{\mathrm{II}}$ produces a graduation rate and a rate of enrollment in high school which are almost identical to those obtained raising school effects on cognitive skills throughout middle school.

The last column of the table shows average outcomes when school effects $\mathcal{R}_{\mathrm{I}}$ on retention are modified to make schools more lenient. This intervention reduces the share of retained students by almost 4 p.p.; the second order effects are small decrease in dropout rate in first period and an increase of the number of students who graduate at time $t = 2$ (the graduation probability at $t = 2$ stays constant and there are more people enrolled in level II), therefore the overall graduation rate increases by 1 p.p. The overall enrollment in high school increases as well by 1 p.p.[42]

average outcomes on the entire sample rather than on the subset of schools affected by each intervention makes results comparable across simulations.

[42]For comparison Table A-20 in the appendix replicates the exercise described in this section using the median rather than the 75 percentile as threshold. Not surprisingly differences across baseline and

Graduation rate and enrollment in high school increase even more for the subsample of students with low parental background, as shown in Table A-21 in the appendix. The table uses the same outputs of Table 12, but frequencies are computed using only individuals with low educated parents. Raising school effects increases graduation rate by up to 6 p.p. and enrollment in high school by up to 6.5 p.p. Next subsection focuses on studying low parental background students and how they are affected by the interventions depending on their school environment.

7.2 Simulations on type L students

I now study what outcomes students with low parental background would have under the various interventions described in previous section. In particular, I analyze how the interventions affect students enrolled in schools with a high or with a low share of other low parental background students. As described in Section 6.1, schools with high share of students with low educated parents are more likely to have higher school effect on choice of staying in school, and slightly more likely to have lower school effect on cognitive skills, but there is large variation in all dimensions. Obviously students enrolled in a school with very large school effects on choices would not be affected by an intervention that raises them, while they may benefit from higher school effects on cognitive skills, and vice-versa for students enrolled in schools with high school effects on cognitive skills.

I simulated outcomes for type L in each school in the sample under the various counterfactuals, following the steps described in Section 6. Table 13 shows predicted outcomes for individuals enrolled in schools with high share of students with low educated parents (panel A), and for individuals enrolled in schools with low share of students with low educated parents (panel B).[43]

Both interventions aimed at raising school effects on cognitive skills and those aimed at improving school effects on choices produce a sizable growth in graduation and high school enrollment rate in all groups of schools. However, in schools with high share of low educated parents targeting effects on cognitive skills has the largest impact, while in schools with low share, targeting effects on choices produces the biggest changes.

In fact, on average raising school effects on cognitive skills at the 75 percentile improves $C_{\mathrm{I},1}$ by 0.28 s.d. in schools with high share of low parental background students, and by 0.14 s.d. in schools with low share. In the former group of schools, both graduation rate and enrollment in high school increase by more than 9 p.p., while raising simultaneously

counterfactuals are smaller, but they follow a similar pattern. The main difference is that the intervention which raises both $\mathcal{T}_{\mathrm{I}}$ and $\mathcal{T}_{\mathrm{II}}$ increases graduation and enrollment in high school more than the intervention which raises $\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$

[43] I ranked schools based on their share of students whose parents attained at most lower secondary education, and grouped them in three quantiles. The 15 schools in the top quantile are used for panel A of Table 13, the 15 schools in the bottom quantile are used for panel B.

the two school effects on tastes increases graduation rate by about 3 p.p. and enrollment in high school by 5.5 p.p. In the latter, raising $\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$ increases graduation rate by 3 p.p. and enrollment in high school by 5 p.p., while raising $\mathcal{T}_{\mathrm{I}}$ and $\mathcal{T}_{\mathrm{II}}$ increases graduation rate by 7.5 p.p. and enrollment in high school by 11 p.p.

Results highlight that similar students may benefit more from interventions targeting the development of their cognitive skills or from interventions targeting non-cognitive skills and their tastes, depending on what the school environment is already offering.

8 Conclusions and future research

Suppose that a policy maker wants to identify the "most successful" institutions, in order to investigate their methodology, learn their best practices, and transfer them to other schools. The school effects on cognitive skills identified through the model presented in this paper or other similar measures of value added would allow her to rank schools based on their capability to improve performance as measured by a nation-wide test. However, it is not evident that attending one of the top performing institutions would be desirable for every type of student, if those institutions do not simultaneously ensure that they help every student to succeed. In fact, graduating from such schools students would potentially reach the highest level of cognitive skills, but this is not happening in practice if they dropout before completing their education. Moreover, in another school they may graduate with a slightly lower level of cognitive skills, but with a stronger motivation to enroll in high school, which may eventually lead to better outcomes in the labor market.

The results presented in this paper confirm that identifying "school quality" with school value added on performances is not a harmless assumption. At most, it might be a viable simplification when focusing on students with favorable socioeconomic conditions, in particular a very high parental background, because they are extremely likely to pursue further academic education no matter what their current school environment is. Conversely, the school environment is a crucial determinant of the educational attainment of students with less advantaged background, not only through its contribution to cognitive skills development, but also above and beyond its effect on their cognitive skills. Evaluating school effectiveness using only performances may lead to conclusions that would not benefit disadvantaged students: a policy maker whose goal is to improve educational outcomes for all should not ignore the other dimensions.

There are several research questions related to the model and the results discussed in this paper that I find appealing. As a follow-up project, I plan to study the long run effects of attending a given middle school on future educational outcomes such as performance and graduation in high school and tertiary education. In particular, I would like to assess whether students who are moved into pursuing further education by an improvement in their non-cognitive skills or in their their consumption value of schooling are able to perform well in the next educational stages and attain higher qualifications. Preliminary results using the additional data which I already collected suggest that they are not more likely to dropout during the first year and perform equally well than similar students from other schools.

Another line of research aims at opening the "black box" of school effects on cognitive skills and educational choice, understanding the mechanisms that lead to the differentiation across schools. A targeted survey administered to principal and teachers would allow me to shed light on how schools differ in teaching methods, remedial and enhancing activities, retention policy, tutoring, inclusion of students from all background, and provision of information on future prospects after graduation. Linking survey data with school effects estimated exploiting administrative data, I would be able to assess what part of the variation across schools can be explained by differing pedagogical approaches and what are the most successful practices to enhance a given outcome among students with given prior characteristics. Moreover, it would allow me to gather information on goals of teachers and administrators (e.g. What do they deem more important, to complete the ministerial curriculum so that students are well prepared for the next educational level if they acquire further education, or to fill the gap for students who are lagging behind even if this may slow down the rest of the class?). In fact schools may optimally set school's commitment to improve students' cognitive skills, their non-cognitive skills and their motivation, in order to maximize an objective function involving students' performance and attainment. On one hand, the finding that schools with a higher share o disadvantaged students are more likely to have large positive effects on the choice of staying in school suggests that schools may adapt to the typical students enrolled there. On the other hand, the large variation of school effects across schools with a similar pool of students suggests that school administrators may differ in their objective functions. The structural model implemented in this paper can consistently estimate the equilibrium results of such decision process, but survey data would provide insights on the process itself. Studying the formation of school effects is important to fully understand the consequences of implementing intervention which redistributed students or resources across schools, given that school administrators may respond to changes adjusting the school policy.

References

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011). Accountability and flexibility in public schools: Evidence from boston's charters and pilots. *The Quarterly Journal of Economics 126*(2), 699.

Allen, C. S., Q. Chen, V. L. Willson, and J. N. Hughes (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis 31*(4), 480–499. PMID: 20717492.

Allen, R. and S. Burgess (2013). Evaluating the provision of school performance information for school choice. *Economics of Education Review 34*(C), 175–190.

Altonji, J. G., P. Arcidiacono, and A. Maurel (2015, October). The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. NBER Working Papers 21655, National Bureau of Economic Research, Inc.

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. R. Walters (2016). Stand and deliver: Effects of bostons charter high schools on college preparation, entry, and choice. *Journal of Labor Economics 34*(2), 275–318.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2017). Leveraging lotteries for school value-added: Testing and estimation*. *The Quarterly Journal of Economics 132*(2), 871–919.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2013, October). Explaining charter school effectiveness. *American Economic Journal: Applied Economics 5*(4), 1–27.

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

Arcidiacono, P., E. Aucejo, A. Maurel, and T. Ransom (2016, June). College attrition and the dynamics of information revelation. Working Paper 22325, National Bureau of Economic Research.

Avery, C., C. Hoxby, C. Jackson, K. Burek, G. Pope, and M. Raman (2006, February). Cost Should Be No Barrier: An Evaluation of the First Year of Harvard's Financial Aid Initiative. NBER Working Papers 12029, National Bureau of Economic Research, Inc.

Belfield, C., T. Boneva, C. Rauh, and J. Shaw (2018). What Drives Enrollment Gaps in Further Education? The Role of Beliefs in Sequential Schooling Decisions. Mimeo.

Bordon, P. and C. Fu (2015). College-major choice to college-then-major choice. *The Review of Economic Studies 82*(4), 1247–1288.

Calsamiglia, C. and A. Loviglio (2018, May). Grading on the curve: when having good peers is not good. Mimeo.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014, September). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Cockx, B., M. Picchio, and S. Baert (2017). Modeling the Effects of Grade Retention in High School. GLO Discussion Paper Series 148, Global Labor Organization (GLO).

Cunha, F. and J. J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources 43*(4).

Cunha, F., J. J. Heckman, and L. Lochner (2006, May). *Interpreting the Evidence on Life Cycle Skill Formation*, Volume 1 of *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Elsevier.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, 05). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Dearden, L., J. Micklewright, and A. Vignoles (2011). The effectiveness of english secondary schools for pupils of different ability levels. *Fiscal Studies 32*(2), 225–244.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw Hill.

Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014, 03). School choice, school quality, and postsecondary attainment. *The American Economic Review 104*(3), 991–1013.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 39*(1), 1–38.

Dobbie, W. and J. Fryer, Roland G. (2011, July). Are high-quality schools enough to increase achievement among the poor? evidence from the harlem children's zone. *American Economic Journal: Applied Economics 3*(3), 158–87.

Dobbie, W. and J. Fryer, Roland G. (2013, October). Getting beneath the veil of effective schools: Evidence from new york city. *American Economic Journal: Applied Economics 5*(4), 28–60.

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

Fryer, Jr., R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics 129*(3), 1355–1407.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2013, July). Are Some Degrees Worth More than Others? Evidence from college admission cutoffs in Chile. NBER Working Papers 19241, National Bureau of Economic Research, Inc.

Heckman, J. J. and Y. Rubinstein (2001, May). The importance of noncognitive skills: Lessons from the ged testing program. *American Economic Review 91*(2), 145–149.

Hoxby, C. M. and C. Avery (2012, December). The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students. NBER Working Papers 18586, National Bureau of Economic Research, Inc.

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on nontest score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

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

James, J. (2011, October). Ability matching and occupational choice. Working Paper 11-25.

Kane, T. J. and D. O. Staiger (2002, December). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives 16*(4), 91–114.

Kinsler, J. and R. Pavan (2015). The specificity of general human capital: Evidence from college major choice. *Journal of Labor Economics 33*(4), 933–972.

Leckie, G. and H. Goldstein (2017). The evolution of school league tables in england 19922016: contextual value-added, expected progress and progress 8. *British Educational Research Journal 43*(2), 193–212.

OECD (2016). Education at a Glance 2016: OECD Indicators. Oecd publishing, Paris.

OECD (2018). Education at a Glance 2016: OECD Indicators. Oecd publishing, Paris.

Pop-Eleches, C. and M. Urquiola (2013, June). Going to a better school: Effects and behavioral responses. *American Economic Review 103*(4), 1289–1324.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica 55*(5), 999–1033.

Wiswall, M. and B. Zafar (2014). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*.

Wiswall, M. and B. Zafar (2015). How Do College Students Respond to Public Information about Earnings? *Journal of Human Capital 9*(2), 117 – 169.

Zafar, B. (2013). College Major Choice and the Gender Gap. *Journal of Human Resources 48*(3), 545–595.

9 Tables

Table 1: Descriptive statistics by subgroups of the population

	N	%	eval. PS	eval. MS	stay at 16	graduate	high school	peers eval.
ALL	5000	1.00	0.00	0.20	0.09	0.83	0.66	-0.00
low parental edu.	1180	0.24	-0.55	-0.32	0.16	0.70	0.42	-0.54
avg parental edu.	1977	0.40	-0.09	0.07	0.10	0.81	0.62	-0.03
high parental edu.	1793	0.36	0.48	0.59	0.02	0.95	0.86	0.40
male	2572	0.51	-0.03	0.22	0.10	0.80	0.60	-0.02
female	2428	0.49	0.03	0.18	0.07	0.86	0.72	0.03
Spanish	4232	0.85	0.11	0.29	0.06	0.87	0.70	0.09
immigrant	768	0.15	-0.61	-0.42	0.23	0.63	0.44	-0.52
regular	3616	0.72	0.28	0.36	0.03	0.97	0.84	0.21
retained in primary	396	0.08	-0.88	-0.77	0.23	0.48	0.26	-0.60
retained in grade 1-3	775	0.15	-0.74	-0.47	0.30	0.40	0.13	-0.60
retained in grade 4	213	0.04	-0.37	-0.39	0.00	0.70	0.27	-0.21

Note. This table report summary statistics for the sample of students used to estimate my structural model. It consists of students who enrolled in a public middle school in Barcelona (Spain) in 2009 or in 2010.

eval. PS are external evaluations at the end of primary school; eval. MS are external evaluations at the end of middle school (computed on the subsample who reached the last grade).

Table 2: Descriptive statistics by schools

	low p.e.	Spanish	eval. PS	retained bfr 15	drop. at 16	graduate	high school	eval. MS
p10	-0.10	0.67	-0.65	0.10	0.03	0.68	0.44	-0.44
p25	-0.16	0.77	-0.35	0.16	0.05	0.73	0.53	-0.16
median	-0.24	0.83	-0.00	0.25	0.10	0.81	0.64	0.14
p75	-0.35	0.91	0.22	0.36	0.16	0.89	0.73	0.39
p90	-0.48	0.95	0.39	0.44	0.18	0.94	0.81	0.53

Note. This table report summary statistics for the 44 public middle schools in Barcelona which are used to estimate my structural model.

Table 3: Variance of unobserved ability

s	$\mu_l^2\widehat{\sigma}$	$\mathrm{Var}(C_l)$	%
$l = 0$	0.278	0.563	49.3
$l = \mathrm{I}$	0.593	1.202	49.3
$l = \mathrm{II}$	0.613	1.047	58.6

Panel A: The first column contains the total variance of cognitive skills before starting middle school, and in level I and II, the second column the variance due to unobserved ability, and the third column the share of total variance due to unobserved ability, i.e. $\mu_l^2\widehat{\sigma}/\mathrm{Var}(C_l)$

Time	Signals received from 0 to t	Posterior Variance
$t = 0$	r_0	0.169
$t = 1$	$r_0, g_{\mathrm{I},1}$	0.0601
$t = 2$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{I},2}$	0.0365
$t = 2$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}$	0.035
$t = 2$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, r_{\mathrm{II},2}$	0.0294
$t = 3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{I},2}, g_{\mathrm{II},3}$	0.0254
$t = 3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{I},2}, g_{\mathrm{II},3}, r_{\mathrm{II},3}$	0.0223
$t = 3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, g_{\mathrm{II},3}$	0.0247
$t = 3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, r_{\mathrm{II},2}, g_{\mathrm{II},3}$	0.0218
$t = 3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, g_{\mathrm{II},3}, r_{\mathrm{II},3}$	0.0218
$t = 3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, r_{\mathrm{II},2}, g_{\mathrm{II},3}, r_{\mathrm{II},3}$	0.0195

Panel B: Posterior variance for each set of signals at a given time

Table 4: Estimates of evaluations parameters

	r_0		g_{I}		r_{II}	g_{II}	
	β_0	β_{I}	γ	$\beta_{\mathrm{I}}+\gamma$	β_{II}	$\mu\beta_{\mathrm{II}}$	$\mu\beta_{\mathrm{II}}+\gamma$
female	0.044	-0.084	0.374	0.290	-0.009	-0.011	0.363
immigrant	-0.265	-0.288	0.116	-0.172	-0.250	-0.294	-0.178
mother education average	0.232	0.250	-0.063	0.187	0.192	0.226	0.164
mother education high	0.471	0.623	-0.088	0.535	0.524	0.617	0.529
father education average	0.235	0.265	-0.041	0.224	0.214	0.252	0.211
father education high	0.377	0.420	-0.013	0.407	0.338	0.397	0.384
day of birth	0.245	0.136	0.008	0.144	0.116	0.136	0.144
retained in primary school	-0.556	-0.709	0.381	-0.328	-0.659	-0.776	-0.395
repeat level	-	0.169	-0.184	-0.014	0.345	0.406	0.223
peers: avg evaluation PS	-	0.178	-0.140	0.037	0.086	0.101	-0.039
peers: avg parental edu	-	0.110	-0.066	0.044	0.017	0.020	-0.046
peers: share of female	-	0.064	-0.025	0.039	-0.002	-0.003	-0.028
peers: share of immigrant	-	-0.030	-0.021	-0.050	-0.027	-0.032	-0.052
peers: share with external	-	0.025	0.006	0.031	-0.023	-0.026	-0.020
peers: share older	-	-0.001	0.011	0.010	0.008	0.009	0.020
school effect p75 - p25	-	0.351	0.219	0.389	0.167	0.196	0.252
school effect p80 - p20	-	0.434	0.289	0.463	0.260	0.306	0.295
previous time-varying regres.	-	0.050	0.000	0.050	0.443	0.522	0.522
unobserved ability	1.000	1.461	-	-	1.263	1.486	-
variance of error	0.434	-	-	0.199	0.292	-	0.185

Note. The estimation includes cohort fixed effects and two dummy variables which take value one if information on mother or father is missing. For each peer regressor a cubic polynomial is used; the table reports the effect on the average student in the sample of having peers 1 s.d. above the mean rather than at the mean. β_{I}, β_{II}, and γ include 44 school effects; the table reports the interquantile range and the difference between the 80 and the 20 percentiles (computed weighting the school effect by size of the school).

I am estimating bootstrap standard errors. They will be added to next version of the working paper.

Table 5: Estimates of retention and graduation parameters

	retention in I	graduation in II
belief on cognitive skills	-1.794	2.477
female	-0.729	0.883
immigrant	0.012	0.178
mother education average	0.012	-0.111
mother education high	-0.283	0.197
father education average	-0.129	0.003
father education high	-0.471	0.233
day of birth	-0.085	0.135
retained in primary school	-	0.584
repeated I	-	0.170
second time in II	-	0.001
peers: avg evaluation PS	0.224	0.248
peers: avg parental edu	-0.007	-0.523
peers: share of female	-0.214	-0.168
peers: share of immigrant	0.005	-0.314
peers: share with external	-1.036	-0.114
peers: share older	0.747	0.088
school effect p75 - p25	0.669	1.361
school effect p80 - p20	0.975	1.863

Note. The estimation includes cohort fixed effects and two dummy variables which take value one if information on mother or father is missing. For each peer regressor a cubic polynomial is used; the table reports the effect on the average student in the sample of having peers 1 s.d. above the mean rather than at the mean. The two logit model include school dummies; the table reports the interquantile range and the difference between the 80 and the 20 percentiles (computed weighting the school effect by size of the school) of the estimated school effects.

I am estimating bootstrap standard errors. They will be added to next version of the working paper.

Table 6: Estimates of choices parameters

	stay in middle school	enroll in high school
$\widehat{C}$ (belief on cognitive skills)	1.395	2.062
$\widehat{C}\times$individual characteristics	0.200	-0.222
$\widehat{C}\times$school effect	0.012	-0.515
$\widehat{C}\times$peers characteristics	0.158	1.014
$\widehat{C}\times$second time in I	-0.408	-
$\widehat{C}\times$second time in II	-0.333	-0.723
$\widehat{C}\times$in II after repeating I	-0.479	-0.574
female	0.353	0.748
immigrant	-0.386	0.621
mother education average	-0.074	0.020
mother education high	0.082	-0.183
father education average	-0.216	0.281
father education high	-0.467	0.615
day of birth	-0.534	-0.158
retained in primary school	0.445	0.944
second time in I	-1.280	-
second time in II	-0.274	-1.624
in II after repeating I	-0.719	-1.705
peers: avg evaluation PS	-0.349	0.011
peers: avg parental edu	0.227	-0.099
peers: share of female	-0.106	0.189
peers: share of immigrant	-0.084	-0.201
peers: share with external	0.211	0.095
peers: share older	-0.033	0.147
school effect p75 - p25	0.669	1.361
school effect p80 - p20	0.975	1.863

Note. The estimation includes cohort fixed effects and two dummy variables which take value one if information on mother or father is missing. For each peer regressor a cubic polynomial is used; the table reports the effect on the average student in the sample of having peers 1 s.d. above the mean rather than at the mean. The model includes school dummies; the table reports the interquantile range and the difference between the 80 and the 20 percentiles (computed weighting the school effect by size of the school) of the estimated school effects.

I am estimating bootstrap standard errors. They will be added to next version of the working paper.

Table 7: Fit of the model

	drop at $t = 1$		graduate		high school		retained in I		retained in II	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
ALL	91.36	91.68	83.20	83.77	65.66	65.63	23.42	22.91	4.26	6.04
male	90.28	90.01	80.09	80.13	60.11	59.67	26.79	26.41	4.82	6.99
female	92.50	93.44	86.49	87.63	71.54	71.94	19.85	19.20	3.67	5.03
Spanish	93.90	93.97	86.81	87.53	69.54	69.29	18.53	18.16	4.11	5.97
immigrant	77.34	79.02	63.28	63.06	44.27	45.46	50.39	49.09	5.08	6.42
low parental edu.	83.90	83.91	69.83	68.42	42.29	41.32	43.47	42.83	6.86	8.28
avg parental edu.	90.19	91.19	81.08	82.60	62.01	62.62	25.70	24.86	4.81	7.32
high parental edu.	97.88	97.54	95.15	95.68	85.78	85.36	6.97	6.87	1.95	3.11
below median peers	86.82	88.23	76.21	77.58	54.00	54.86	32.78	31.55	4.95	6.28
above median peers	94.65	94.18	88.27	88.27	74.12	73.44	16.63	16.64	3.76	5.86

Note. Data frequencies are computed on the sample of students used in the estimation. The model frequencies are constructed using 100 simulations of the structural model for each individual included in the estimation.

Table 8: Average school environment by student type

	% female	% immigrant	avg p.e.	% with eval. PS	avg eval. PS	% older peers
low parental edu.	0.487	0.272	0.362	0.795	0.720	0.221
(percentile)	46	75	30	24	23	73
high parental edu.	0.494	0.116	0.581	0.888	0.802	0.098
(percentile)	47	40	69	57	64	45

Panel A: Peers' characteristics

	$\mathcal{I}$	$\mathcal{A}_{\text{I}}$	$\mathcal{A}_{\text{II}}$	$\mathcal{T}_{\text{I}}$	$\mathcal{T}_{\text{II}}$	$\mathcal{R}_{\text{I}}$	$\mathcal{G}_{\text{II}}$
low parental edu.	-0.167	0.530	0.466	-0.450	-0.159	0.339	-0.410
(percentile)	55	43	52	57	42	43	45
high parental edu.	-0.211	0.600	0.530	-0.767	-0.198	0.643	-0.327
(percentile)	42	59	58	37	39	59	53

Panel B: School dummy coefficients

Table 9: Educational outcomes by student type and environment

	Avg. p.e.	Low parental education				High parental education			
	(1)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
repeat level I	0.283	0.594	0.633	0.302	0.337	0.041	0.035	0.134	0.115
dropout at t=1	0.056	0.147	0.198	0.067	0.093	0.007	0.006	0.023	0.017
drop at t=2|not grad.	0.141	0.230	0.274	0.152	0.185	0.049	0.039	0.095	0.079
graduate	0.891	0.687	0.630	0.836	0.808	0.990	0.992	0.964	0.973
enrol in high school	0.652	0.248	0.238	0.434	0.435	0.937	0.935	0.868	0.871
enrol in h.s.|grad.	0.731	0.362	0.377	0.519	0.538	0.947	0.942	0.901	0.895
r_0	0.063	-0.349	-0.349	-0.349	-0.349	0.499	0.499	0.499	0.499
$C_{\mathrm{I},1}$	0.109	-0.574	-0.504	-0.251	-0.181	0.862	0.792	0.540	0.469
$g_{\mathrm{I},1}$	-0.035	-0.467	-0.440	-0.425	-0.398	0.544	0.517	0.502	0.475
r_{II}	0.195	-0.210	-0.154	-0.103	-0.048	0.696	0.640	0.533	0.474
g_{II}	-0.068	-0.347	-0.325	-0.435	-0.414	0.416	0.395	0.460	0.437

Note. Frequencies are constructed using 10000 simulations of the structural model for each student type. Type L has parents with primary education; type H has parents with tertiary education. For comparison the first column of the table reports outcomes for a student with the average parental characteristics in the sample. The fictitious students created in the simulation are male, Spanish, born on July 1, they began middle school at 12 years old and they are assigned the average cohort effect and the average primary school effect in the sample. Columns (1) contain results of the baseline specification in which each peers characteristic and school effect takes the average values among students with the same parental background. In columns (2) average school effects among students with highly educated parents are used for type L, and vice-versa average school effects among students with low educated parents are used for type H. In columns (3) average peers characteristics of the opposite type are used; in columns (4) both school effects and peers of the other type are used.

Table 10: Educational outcomes by schools

	Low parental education				High parental education			
	p50	p80-p20	p75-p25	p70-p30	p50	p80-p20	p75-p25	p70-p30
repeat level I	0.451	0.303	0.236	0.191	0.071	0.091	0.068	0.051
dropout at t=1	0.111	0.098	0.080	0.054	0.009	0.016	0.012	0.009
drop. at t=2|do not grad.	0.206	0.097	0.079	0.060	0.051	0.043	0.040	0.034
graduate	0.744	0.204	0.184	0.134	0.986	0.027	0.021	0.016
enrol in high school	0.323	0.258	0.177	0.128	0.906	0.091	0.073	0.053
enrol in h.s.|grad.	0.437	0.228	0.184	0.131	0.921	0.079	0.052	0.036
$C_{\mathrm{I},1}$	-0.355	0.655	0.519	0.407	0.688	0.655	0.519	0.407
C_{II}	-0.323	0.611	0.446	0.388	0.657	0.589	0.463	0.353

Note. Frequencies are constructed using 10000 simulations of the structural model for each type and every school of the sample. Type L has parents with primary education; type H has parents with tertiary education. For comparison the first column of the table reports outcomes for a student with the average parental characteristics in the sample. The fictitious students created in the simulation are male, Spanish, born on July 1, they began middle school at 12 years old and they are assigned the average cohort effect and the average primary school effect in the sample. The average peers characteristics in each school are used.

Table 11: Educational outcomes by retention status. Student with low educated parents

	not retained	retained in I	retained in II
dropout at t=1	0.1025	0.1759	-
graduate (Pr at t=1)	0.8691	0.5621	-
enrol in high school (Pr at t=1)	0.4208	0.1288	-
graduate at t=2|stay at t=1	0.8593	-	-
enrol in hs at t=2|grad. at t=2	0.5074	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.2379	0.1623
graduate at t=3|stay at t=2	-	0.8951	0.9253
enrol in hs at t=3|grad. at t=3	-	0.2291	0.3013
true $C_{\mathrm{I},1}$	-0.5741	-0.5741	-
perceived $\widehat{C}_{\mathrm{I},1}$	-0.5246	-0.6049	-
true $C_{\mathrm{I},2}$	-	-0.4046	-
perceived $\widehat{C}_{\mathrm{I},2}$	-	-0.4064	-
true $C_{\mathrm{II},2}$	-0.5000	-	-0.5000
perceived $\widehat{C}_{\mathrm{II},2}$	-0.4507	-	-0.5696
true C_{II}|graduate	-0.5000	-0.4116	-0.0939
perceived $\widehat{C}_{\mathrm{II}}$|graduate	-0.4507	-0.3892	-0.1270

Note. Frequencies are constructed using 10000 simulations of the structural model. The following characteristics are used: parents with primary education, male, Spanish, born on July 1, began middle school in the year in which they turn 12. The average cohort effect and the average primary school effect in the sample are used. Peers characteristics and school effects are the average values among students with low educated parents.

Table 12: Simulated outcomes for the entire sample. School effects raised at 75 percentile.

	Baseline	Counterfactuals				
school effects modified		$\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}$	$\mathcal{T}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}}$	$\mathcal{R}_{\mathrm{I}}$
repeat level I	0.230	0.202	0.230	0.230	0.230	0.193
dropout at t=1	0.083	0.064	0.061	0.076	0.056	0.079
grad at t=2| in II	0.918	0.931	0.916	0.918	0.915	0.913
drop. at t=2|not grad.	0.212	0.192	0.177	0.196	0.164	0.215
grad at t=3| in II	0.847	0.860	0.829	0.846	0.828	0.840
graduate	0.838	0.870	0.858	0.846	0.864	0.847
enrol in high school	0.657	0.702	0.666	0.698	0.708	0.666
enrol in hs|graduation	0.784	0.807	0.776	0.825	0.819	0.786
$C_{\mathrm{I},1}$	0.001	0.192	0.001	0.001	0.001	0.001
C_{II}|graduation	0.264	0.359	0.247	0.258	0.242	0.258

Note. Average outcomes in the column "baseline" are computed using the estimated parameters of the model. In the counterfactuals, the school effects reported above each column are replaced with values at the 75 percentile if they are lower. Frequencies are constructed using 100 simulations of the structural model for each individual included in the estimation.

Table 13: Simulated outcomes for type L. School effects raised at 75 percentile.

Panel A: Schools with high share of students with low educated parents

	Baseline	Counterfactuals				
school effects modified		$\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}$	$\mathcal{T}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}}$	$\mathcal{R}_{\mathrm{I}}$
repeat level I	0.575	0.483	0.575	0.575	0.575	0.514
dropout at t=1	0.145	0.100	0.121	0.139	0.117	0.140
grad at t=2| in II	0.848	0.881	0.845	0.847	0.845	0.833
drop. at t=2|not grad.	0.242	0.201	0.216	0.232	0.210	0.240
grad at t=3| in II	0.861	0.894	0.854	0.861	0.854	0.863
graduate	0.672	0.766	0.699	0.682	0.705	0.691
enrol in high school	0.216	0.308	0.223	0.261	0.270	0.230
enrol in hs|graduation	0.321	0.402	0.319	0.383	0.382	0.333
$C_{\mathrm{I},1}$	-0.784	-0.526	-0.784	-0.784	-0.784	-0.784
C_{II}|graduation	-0.585	-0.426	-0.579	-0.582	-0.579	-0.580

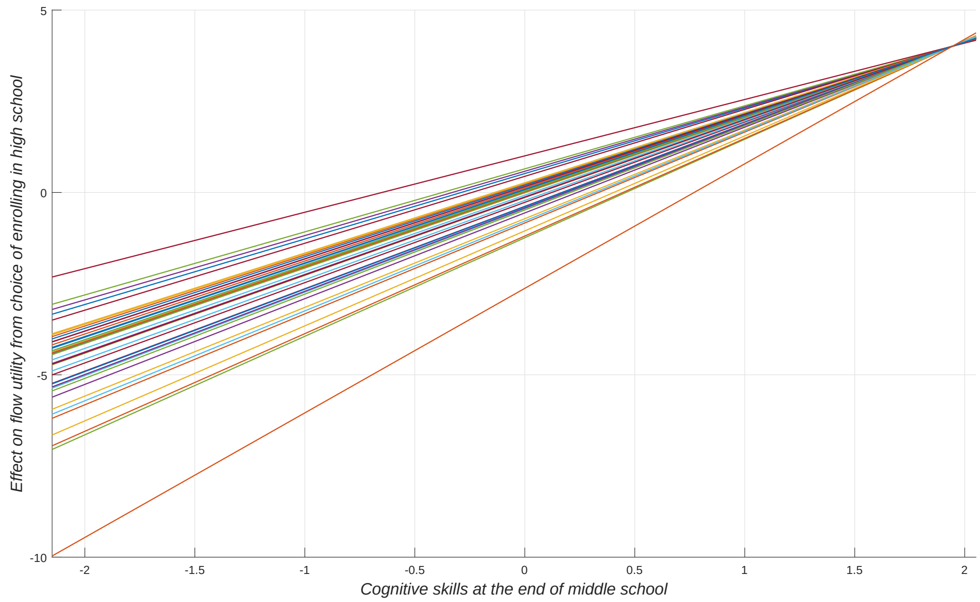
Panel B: Schools with low share of students with low educated parents

	Baseline	Counterfactuals				
school effects modified		$\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}$	$\mathcal{T}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}}$	$\mathcal{R}_{\mathrm{I}}$
repeat level I	0.366	0.322	0.366	0.366	0.366	0.229
dropout at t=1	0.103	0.083	0.047	0.095	0.042	0.091
grad at t=2| in II	0.855	0.873	0.853	0.854	0.853	0.844
drop. at t=2|not grad.	0.214	0.203	0.135	0.192	0.119	0.215
grad at t=3| in II	0.885	0.903	0.867	0.882	0.865	0.897
graduate	0.779	0.818	0.846	0.793	0.855	0.821
enrol in high school	0.421	0.475	0.448	0.503	0.537	0.469
enrol in hs|graduation	0.540	0.580	0.530	0.634	0.628	0.571
$C_{\mathrm{I},1}$	-0.101	0.029	-0.101	-0.101	-0.101	-0.101
C_{II}|graduation	-0.077	-0.005	-0.067	-0.072	-0.064	-0.072

Note. Average outcomes in the column "baseline" are computed using the estimated parameters of the model. In the counterfactuals, the school effects reported above each column are replaced with values at the 75 percentile if they are lower. Frequencies are constructed using 10000 simulations of the structural model for each school of the type L student described in Section 6. Panel A shows average outcomes among individual enrolled in schools with high share of students with low educated parents (the top third of schools); panel B shows average outcomes for schools with low share of students with low educated parents (the bottom third).

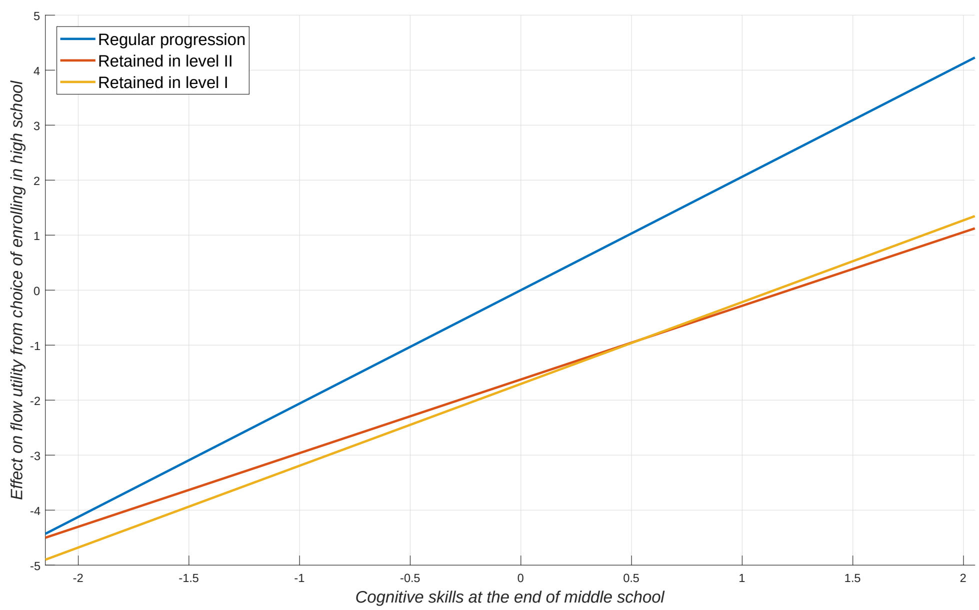
10 Figures

Figure 1: Effect of cognitive skills on high school enrollment by middle school attended



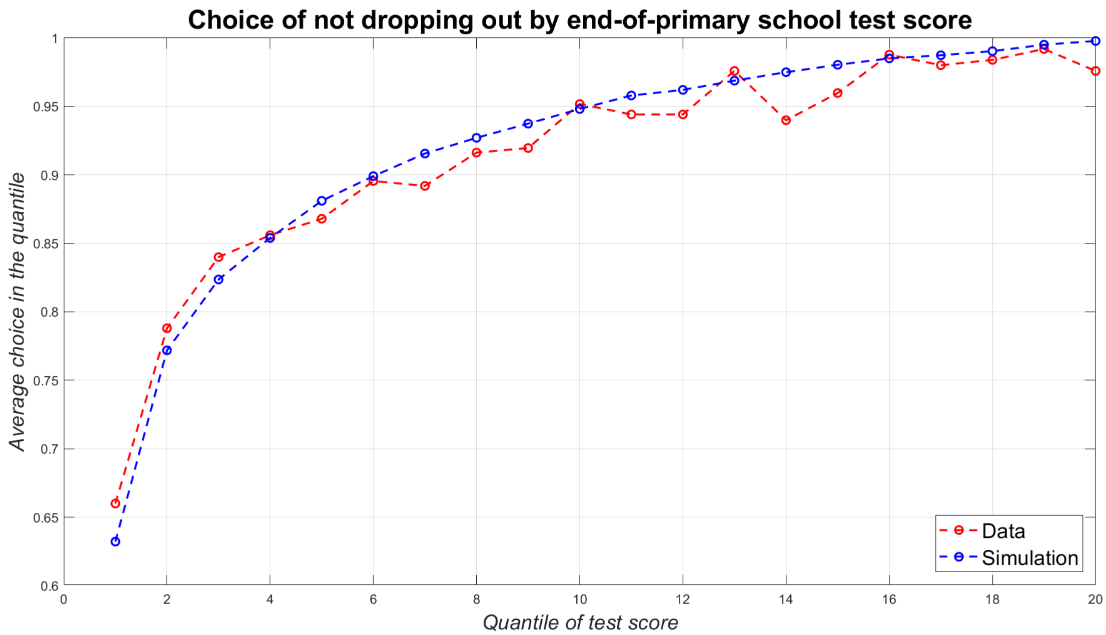
Note. The figure plots the effect of beliefs about cognitive skills on the flow utility of the choice of enrolling in high school after graduation. Each line represents a school in the sample.

Figure 2: Effect of cognitive skills on high school enrollment by retention status



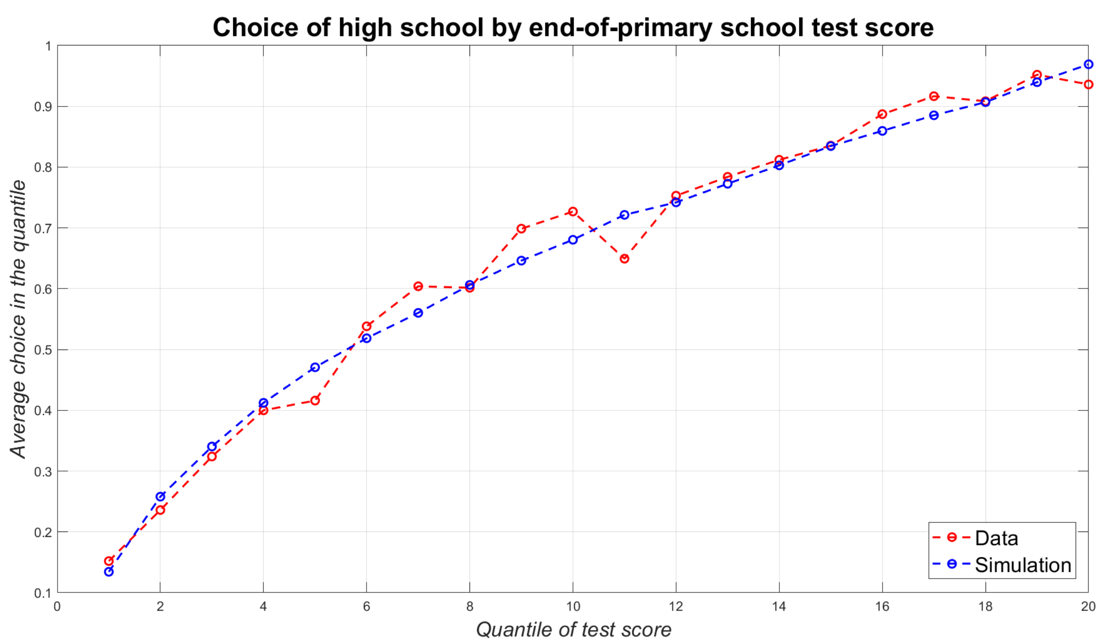
Note. The figure plots the effect of beliefs about cognitive skills on the utility from the choice of enrolling in high school for three types of students: those who completed middle school regularly at time $t = 2$ (blue line), those who graduated at time $t = 3$ because they were retained at time$t = 2$ and repeat level II (red line), those who graduated at time $t = 3$ because they were retained at time$t = 1$ and repeat level I (orange line).

Figure 3: Fit of the model (i)



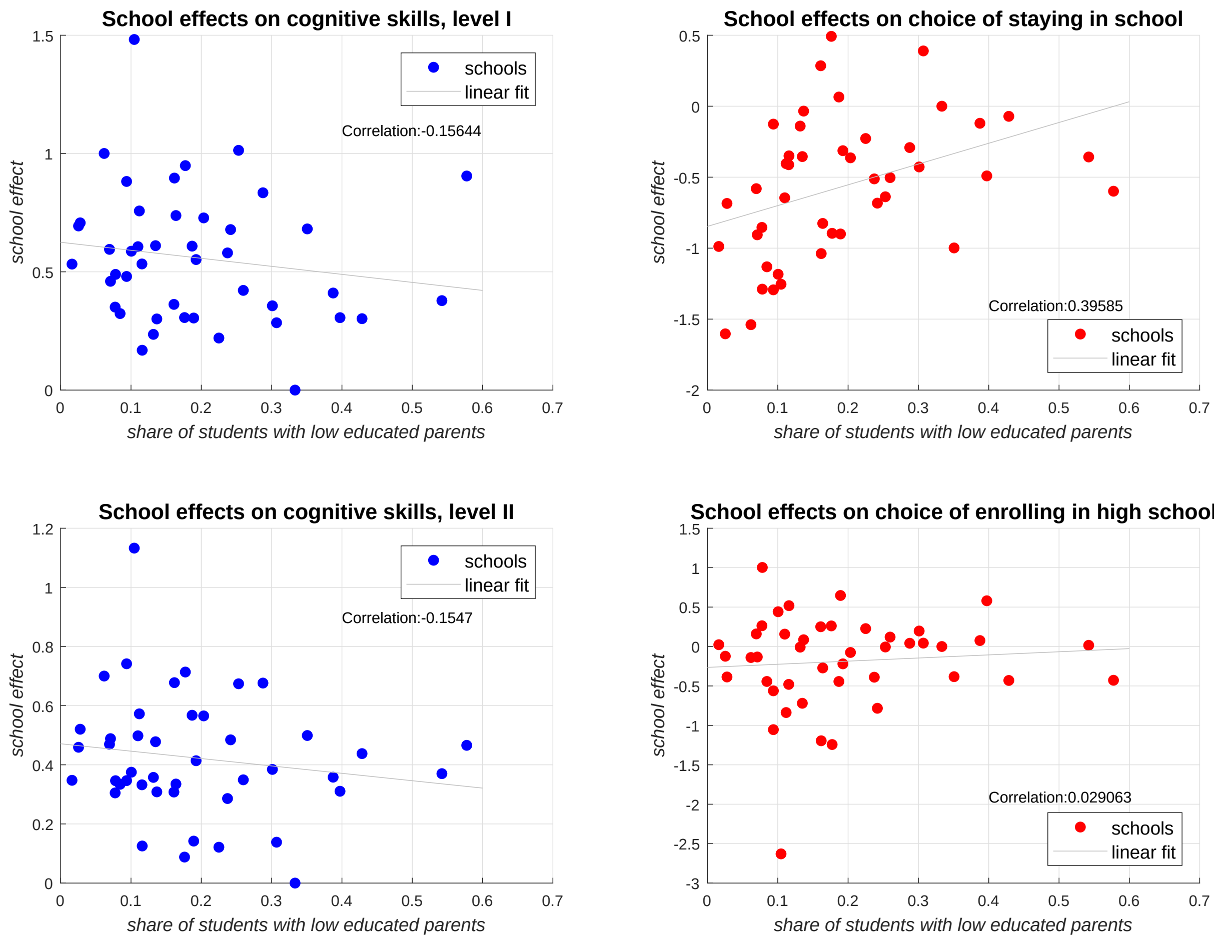
Note. The figure plots the share of students who chose to stay in school at time $t = 1$ by quantile of their test score at the end of primary school. Sample average from the real data are in red, while results of the simulation performed using the estimated parameters of the model are in blue.

Figure 4: Fit of the model (ii)



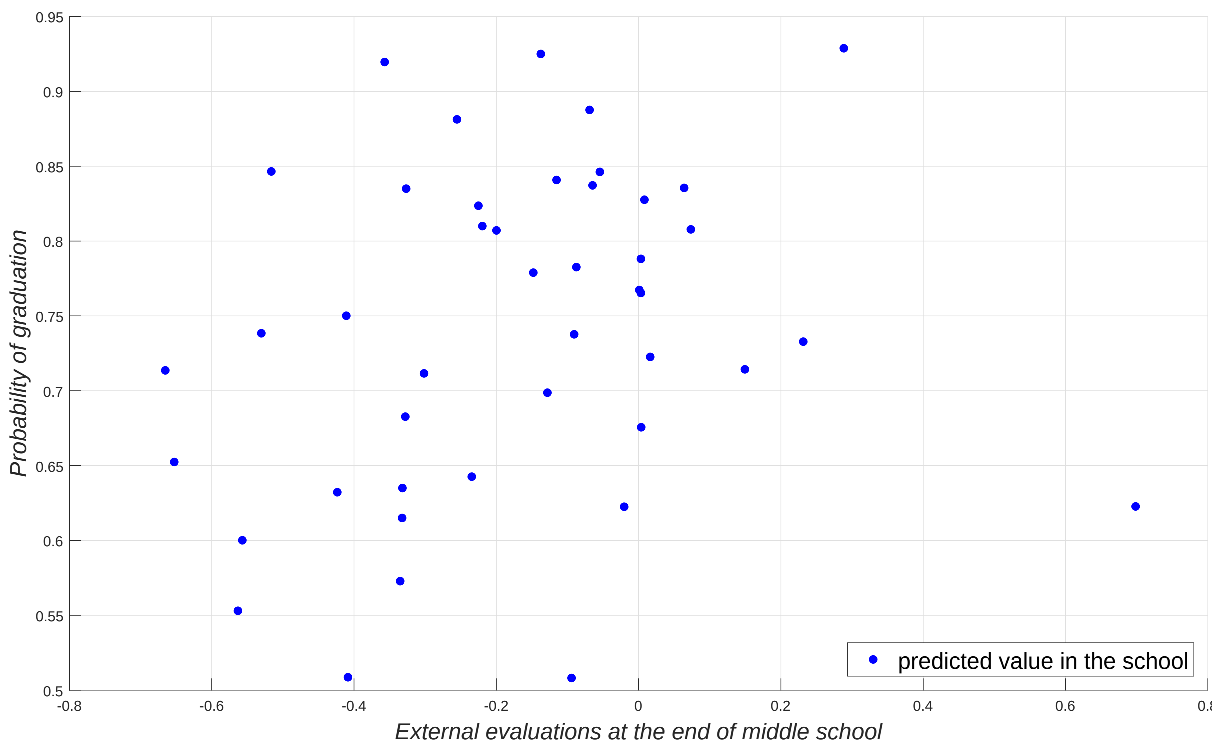
Note. The figure plots the share of students who enroll in high school by quantile of their test score at the end of primary school. Sample average from the real data are in red, while results of the simulation performed using the estimated parameters of the model are in blue.

Figure 5: Share of students with low educated parents *vs* school effects



Note. The figure plots the estimated school effects against the share of students with low educated parents (at most lower secondary education) enrolled in the school. The left side plots school effects on cognitive skills, the right side plots school effects on choices.

Figure 6: Probability of graduation *vs* predicted final evaluations by middle school attended



Note. The figure plots predicted outcomes for a Spanish male student whose parents are low educated (at most lower secondary education).

A Additional tables

Table A-14: Fit of the model (bis)

	graduate at $t = 2$		graduate at $t = 3$		enroll in hs	
	Data	Model	Data	Model	Data	Model
ALL	94.28	91.84	81.40	84.62	78.92	78.34
male	93.23	90.03	78.62	82.39	75.05	74.47
female	95.29	93.56	84.94	87.54	82.71	82.10
Spanish	94.87	92.45	81.57	85.48	80.10	79.16
immigrant	88.15	86.06	80.99	81.79	69.96	72.09
low parental edu.	87.20	84.20	83.18	81.08	60.56	60.38
avg parental edu.	93.26	89.78	80.50	85.19	76.48	75.82
high parental edu.	97.89	96.60	82.52	92.48	90.15	89.22
below median peers	92.26	90.26	82.09	84.00	70.85	70.72
above median peers	95.41	92.75	80.56	85.28	83.97	83.20

Note. Statistics are conditional on reaching the relevant level

Table A-15: Educational outcomes by retention status. Student with low educated parents. Known ability.

	not retained	retained in I	retained in II
dropout at t=1	0.1072	0.1626	-
graduate (Pr at t=1)	0.8698	0.5771	-
enrol in high school (Pr at t=1)	0.4001	0.1253	-
graduate at t=2|stay at t=1	0.8644	-	-
enrol in hs at t=2|grad. at t=2	0.4800	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.2366	0.1426
graduate at t=3|stay at t=2	-	0.9027	0.9441
enrol in hs at t=3|grad. at t=3	-	0.2171	0.3024
true $C_{\mathrm{I},1}$	-0.5741	-0.5741	-
true $C_{\mathrm{I},2}$	-	-0.4046	-
true $C_{\mathrm{II},2}$	-0.5000	-	-0.5000
true C_{II}|graduate	-0.5000	-0.4116	-0.0939

Table A-16: Educational outcomes by student type and environment - with known ability

	Avg p.e.	Low parental education				High parental education			
	(1)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
repeat level I	0.262	0.602	0.644	0.281	0.320	0.032	0.026	0.114	0.098
dropout at t=1	0.043	0.141	0.192	0.060	0.083	0.004	0.004	0.014	0.011
graduate	0.912	0.694	0.633	0.855	0.827	0.994	0.995	0.975	0.981
enrol in high school	0.673	0.235	0.222	0.439	0.440	0.946	0.943	0.886	0.885
enrol in hs|graduation	0.738	0.338	0.351	0.513	0.531	0.952	0.948	0.908	0.902

Table A-17: Educational outcomes by retention status. Student with high educated parents

	not retained	retained in I	retained in II
dropout at t=1	0.0056	0.0333	-
graduate (Pr at t=1)	0.9938	0.8956	-
enrol in high school (Pr at t=1)	0.9472	0.6842	-
graduate at t=2|stay at t=1	0.9895	-	-
enrol in hs at t=2|grad. at t=2	0.9548	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.0658	0.0489
graduate at t=3|stay at t=2	-	0.9917	0.9979
enrol in hs at t=3|grad. at t=3	-	0.7639	0.7826
true $C_{\mathrm{I},1}$	0.8624	0.8624	-
perceived $\widehat{C}_{\mathrm{I},1}$	0.8674	0.7857	-
true $C_{\mathrm{I},2}$	-	1.0319	-
perceived $\widehat{C}_{\mathrm{I},2}$	-	0.9901	-
true $C_{\mathrm{II},2}$	0.8076	-	0.8076
perceived $\widehat{C}_{\mathrm{II},2}$	0.8123	-	0.6819
true C_{II}|graduate	0.8076	0.8960	1.2136
perceived $\widehat{C}_{\mathrm{II}}$|graduate	0.8123	0.8768	1.1301

Table A-18: Educational outcomes by school effects

	Low parental education				High parental education			
	p50	p80-p20	p75-p25	p70-p30	p50	p80-p20	p75-p25	p70-p30
repeat level I	0.471	0.255	0.225	0.180	0.075	0.081	0.070	0.054
dropout at t=1	0.110	0.103	0.082	0.039	0.011	0.011	0.009	0.006
drop. at t=2|do not grad.	0.203	0.114	0.078	0.045	0.052	0.042	0.034	0.027
graduate	0.736	0.189	0.146	0.093	0.984	0.019	0.014	0.009
enrol in high school	0.318	0.167	0.133	0.103	0.913	0.053	0.048	0.033
enrol in h.s.|grad.	0.449	0.185	0.146	0.127	0.927	0.047	0.035	0.028
$C_{\mathrm{I},1}$	-0.394	0.445	0.380	0.325	0.650	0.445	0.380	0.325
C_{II}	-0.341	0.317	0.283	0.201	0.632	0.306	0.238	0.179

Note. Frequencies are constructed using 10000 simulations of the structural model for each type and every school of the sample. Type L has parents with primary education; type H has parents with tertiary education. For comparison the first column of the table reports outcomes for a student with the average parental characteristics in the sample. The fictitious students created in the simulation are male, Spanish, born on July 1, they began middle school at 12 years old and they are assigned the average cohort effect and the average primary school effect in the sample. The average peers characteristics in the sample are used.

Table A-19: Educational outcomes by peers at school

	Low parental education				High parental education			
	p50	p80-p20	p75-p25	p70-p30	p50	p80-p20	p75-p25	p70-p30
repeat level I	0.463	0.364	0.322	0.247	0.074	0.113	0.100	0.076
dropout at t=1	0.130	0.103	0.089	0.062	0.012	0.018	0.013	0.009
drop. at t=2|do not grad.	0.218	0.101	0.084	0.059	0.057	0.057	0.042	0.023
graduate	0.734	0.188	0.171	0.154	0.983	0.031	0.023	0.015
enrol in high school	0.324	0.217	0.188	0.126	0.906	0.091	0.079	0.041
enrol in h.s.|grad.	0.433	0.164	0.136	0.110	0.921	0.070	0.052	0.032
$C_{\mathrm{I},1}$	-0.374	0.339	0.303	0.271	0.669	0.339	0.303	0.271
C_{II}	-0.317	0.231	0.219	0.186	0.653	0.227	0.193	0.180

Note. Frequencies are constructed using 10000 simulations of the structural model for each type and every school of the sample. Type L has parents with primary education; type H has parents with tertiary education. For comparison the first column of the table reports outcomes for a student with the average parental characteristics in the sample. The fictitious students created in the simulation are male, Spanish, born on July 1, they began middle school at 12 years old and they are assigned the average cohort effect and the average primary school effect in the sample. The average peers characteristics in each school are used, while school fixed effects are set at their average value in the sample.

Table A-20: Simulated outcomes for the entire sample. School effects raised at the median.

school effects modified	Baseline	Counterfactuals $\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}$	$\mathcal{T}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}}$	$\mathcal{R}_{\mathrm{I}}$
repeat level I	0.230	0.215	0.230	0.230	0.230	0.211
dropout at t=1	0.083	0.073	0.070	0.079	0.056	0.081
grad at t=2| in II	0.918	0.924	0.917	0.918	0.915	0.915
drop. at t=2|not grad.	0.212	0.203	0.193	0.203	0.164	0.213
grad at t=3| in II	0.847	0.852	0.837	0.846	0.828	0.844
graduate	0.838	0.854	0.849	0.843	0.864	0.843
enrol in high school	0.657	0.677	0.662	0.680	0.708	0.662
enrol in hs|graduation	0.784	0.793	0.780	0.807	0.819	0.785
$C_{\mathrm{I},1}$	0.001	0.092	0.001	0.001	0.001	0.001
C_{II}|graduation	0.264	0.297	0.256	0.262	0.242	0.261

Note. Average outcomes in the column "baseline" are computed using the estimated parameters of the model. In the counterfactuals, the school effects reported above each column are replaced with values at the median if they are lower. Frequencies are constructed using 100 simulations of the structural model for each individual included in the estimation.

Table A-21: Simulated outcomes on the subsample of students with low educated parents.

school effects modified	Baseline	Counterfactuals $\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}$	$\mathcal{T}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}}$	$\mathcal{R}_{\mathrm{I}}$
repeat level I	0.424	0.376	0.424	0.424	0.424	0.374
dropout at t=1	0.163	0.126	0.128	0.153	0.122	0.158
grad at t=2| in II	0.848	0.872	0.843	0.846	0.842	0.838
drop. at t=2|not grad.	0.248	0.220	0.217	0.233	0.203	0.249
grad at t=3| in II	0.818	0.833	0.799	0.817	0.799	0.813
graduate	0.687	0.745	0.715	0.699	0.725	0.700
enrol in high school	0.402	0.465	0.411	0.458	0.470	0.413
enrol in hs|graduation	0.585	0.623	0.576	0.655	0.649	0.590
$C_{\mathrm{I},1}$	-0.810	-0.595	-0.810	-0.810	-0.810	-0.810
C_{II}|graduation	-0.359	-0.269	-0.373	-0.363	-0.378	-0.363

Note. Average outcomes in the column "baseline" are computed using the estimated parameters of the model. In the counterfactuals, the school effects reported above each column are replaced with values at the 75 percentile if they are lower. This table exploits results of the simulations presented in Table 12, but frequencies are computed using only the subsample of students with low educated parents (both mother and father have at most lower secondary education).

B EM algorithm: theoretical framework

Let ζ be the vector of all the parameters that enter the grades equations (including variances of the errors); recall that σ is the variance of the ability h. The likelihood $L(o_i;\zeta,\sigma)$ is the joint density function of the outcomes. As discussed in previous section

$$\log L(o_i;\zeta,\sigma) = \log \int L(o_i;\zeta,\sigma|h)\boldsymbol{\phi}(h)\,dh \tag{A-1}$$

$$L(o_i;\zeta,\sigma|h) = L(r_{i,0};\zeta,\sigma|h)L(g_{i,1};\zeta,\sigma|h)...L(o_{i,T_d};\zeta,\sigma|h) \tag{A-2}$$

where the likelihood of each evaluation conditional on h is a normal density function. For instance:

$$L(r_{i0};\zeta,\sigma|h) = \frac{1}{\sqrt{2\pi\rho_0^r}}\exp\left(-\frac{(r_{i,0}-h-z'_{i,0}\beta_0)^2}{2\rho_0^r}\right) \tag{A-3}$$

Taking the log of (A-2) would simplify the expression and allow an easy estimation through maximum likelihood. Unfortunately the integral over h prevent us from doing so. The proposed approach aims at overcoming this issue.

The FOC of the sum of individual log-likelihoods are as follow:

$$\frac{\partial}{\partial\zeta}\sum_i \log L(o_i;\zeta,\sigma) = \sum_i \frac{1}{L(o_i;\zeta,\sigma)}\int \frac{\partial L(o_i;\zeta,\sigma|h)}{\partial\zeta}\boldsymbol{\phi}(h)\,dh = 0 \tag{A-4}$$

$\boldsymbol{\psi}_i(h) = \boldsymbol{\psi}(h|o_i;\zeta,\sigma)$ is the conditional density of h for individual i given her outcomes and the parameters. By definition of conditional density

$$\boldsymbol{\psi}_i(h) = \frac{L(o_i;\zeta,\sigma|h)\boldsymbol{\phi}(h)}{L(o_i;\zeta,\sigma)} \tag{A-5}$$

Now, moving $L(o_i;\zeta,\sigma)$ under the integral and multiplying by $1 = \frac{L(o_i;\zeta,\sigma|h)}{L(o_i;\zeta,\sigma|h)}$, equation (A-4) can be rewritten as

$$\sum_i \int \frac{L(o_i;\zeta,\sigma|h)\boldsymbol{\phi}(h)}{L(o_i;\zeta,\sigma)}\frac{1}{L(o_i;\zeta,\sigma|h)}\frac{\partial L(o_i;\zeta,\sigma|h)}{\partial\zeta}\,dh = \tag{A-6}$$

$$=\sum_i \int \frac{1}{L(o_i;\zeta,\sigma|h)}\frac{\partial L(o_i;\zeta,\sigma|h)}{\partial\zeta}\boldsymbol{\psi}_i(h)\,dh = \sum_i \int \frac{\partial}{\partial\zeta}\Big(\log L(o_i;\zeta,\sigma|h)\Big)\boldsymbol{\psi}_i(h)\,dh = \tag{A-7}$$

$$=\frac{\partial}{\partial\zeta}\left[\sum_i \int \log L(o_i;\zeta,\sigma|h)\boldsymbol{\psi}_i(h)\,dh\right] = 0 \tag{A-8}$$

Thus if $\widehat{\zeta}$ solves equation (A-4) it solves also equation (A-8) and vice-versa. The advantage of the second object is that it allows to work with $\log L(o_i; \zeta, \sigma|h)$ and the individual posterior distributions. In next section I will give an explicit formulation for it.

Parameters can be estimated using an iterative algorithm which is a taylored application of the EM algorithm. In a nutshell, at each iteration k, first (E-step) posterior distributions $\boldsymbol{\psi}_i^k(h)$ are estimated for all individuals using previous iteration estimates ζ^{k-1} . Then (M-step) estimates of pararameters ζ^k are computed as solution of

$$\zeta^k = \arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma^k|h) \boldsymbol{\psi}_i^k(h)\, dh \tag{A-9}$$

The general theory ensures convergence of the algorithm.[44]

[44] Dempster, Laird, and Rubin (1977)