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

Annalisa Loviglio*

August 10, 2020

(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 effects 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 that increase educational attainment, not only test scores.

JEL codes: C35, I24, J24.

*Department of Economics, University of Bologna, Italy. Correspondence: annalisa.loviglio@unibo.it.
I am grateful to my PhD advisors Caterina Calsamiglia and Joan Llull for their guidance and continuous support. I would also like to thank Manuel Arellano, Peter Arcidiacono, Chiara Binelli, Massimilano Bratti, Ana Rute Cardoso, Olivier De Groote, Anna Houstecka, Arnaud Maurel, Lavinia Piemontese, and participants at the Applied Working Group at Universitat Autònoma de Barcelona 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 that raises outcomes for all, particularly the most disadvantaged, is a long standing preoccupation for policy makers all around the world. Many countries have implemented 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. I exploit administrative data from public schools in Barcelona (Spain), a setting in which 17% of students do not complete basic education and 35% of students do not enroll in high school. 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

[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).

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. 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 "academic track" that 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. Moreover, the school environment that the student experienced may substantially affect her desire to acquire further education.

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 unknown ability, individual characteristics, and school environment (captured by peer characteristics and school inputs). 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'

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

[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.

educational choices in different ways. 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 since they often receive more negative signals.

I estimate this model following an approach that 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 in the years 2009-2015. In this setting, nationwide 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 several 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 grading policies vary 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 the school environment is an important determinant of cognitive skills, both through peer quality (being with 1 standard deviation higher quality peers increases cognitive skills by up to 0.3 standard deviation) and through the school inputs beyond peers (the interquantile range of school inputs is about 0.4 s.d.). Moreover the 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 utility from pursuing further education about as much as increasing cognitive skills by 0.5 s.d.).

I use the model to perform several simulation exercises. In one of them, I assess 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 compare outcomes of students with identical individual characteristics and innate ability but differing parental background: the parents of one of them are highly educated, while the others are low educated. For both students, the school environment is an important determinant of cognitive skills development. However for the student with highly educated parents the middle school attended has relatively little importance for his educational attainment, because he is extremely likely to graduate and enroll in upper secondary education regardless of the school environment. Conversely, for the student with low educated parents the middle school attended plays a crucial role: in some schools he graduates almost for sure while in others he has one chance out of three to leave education without a basic qualification. Similarly, after graduating from some middle schools he has 70% probability of choosing to enroll in upper secondary education, while in others the enrollment probability is only 40%. Interestingly, the results of the simulation show that the correlation between expected performance and probability of graduation or enrollment in upper secondary education is not large: disadvantaged students have better chances to graduate and pursue further studies in several schools with average predicted evaluations than in other schools with a potentially higher final test score.

In another exercise, I evaluate the impact of a change in law that increase compulsory education and the impact of interventions that improve selected school inputs. Raising compulsory schooling substantially improve graduation rate. In fact, most students who would have drop out have the necessary skills to pass the final evaluations and successfully graduate if they stay in school. Both interventions that raise school inputs on cognitive skills and interventions that raise school inputs on choices largely increase enrollment in upper secondary education. Students with low parental background are those who benefit the most from interventions that do not only target cognitive skills.

I also investigate 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. Schools vary widely in their retention policies, therefore their leniency or strictness has a relevant impact on students' attainment. Retained students have somewhat lower beliefs about their ability than otherwise identical students promoted to next level. However a counterfactual simulation without uncertainty confirms that most of the differences in their choices is due to changes in preferences, rather than to misperception of their ability.

Related Literature

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 and 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 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 attainment that 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

[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.

[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]Among recent examples, Belfield, Boneva, Rauh, and Shaw (2018) 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. Giustinelli and Pavoni (2017) study high school track choice in Italy and find that children from less advantaged families display lower initial perceived knowledge and acquire information at a slower pace.

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 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.[9]

I find that the school environment is particularly relevant for choices and attainment 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 if performance of all students are affected similarly by a given school environment, schools may matter differently for educational attainment 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.[10] Interestingly, the gap is larger for

[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.

[9] Beuermann, Jackson, Navarro-Sola, and Pardo (2019), written contemporaneously to this paper, study schools in Trinidad and Tobago and show that schools' impacts on high-stakes tests are weakly related to other life outcomes such as crime or teen motherhood. Their finding also suggest that evaluations based solely on test scores may be very misleading about school quality.

[10] 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).

students with relatively higher cognitive skills, 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 discusses simulations and counterfactual analysis, with a focus on outcomes of students with low parental background. Section 7 concludes.

2 Data

I employ administrative data on the universe of students who began lower secondary education in one of 47 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. By law, they can be retained at most once, thus they start middle school either in the year in which they turn 12 or one year after. In middle school students can be retained at most twice, in two different grades. Thus, they can graduate in the year in which they turn 16 years old or later on, depending on their previous retention history.[11] Repeating a grade during lower secondary is fairly common, while repeating twice is very rare. In public schools in Barcelona about 20% of students are retained once, but only 3% repeat two grades.

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.

[11] For primary education: Decret 142/2007, issued on June, 26 (in DOGC núm. 4915 - 26/6/2007). For secondary education: Decret 143/2007, issued on June, 26 (in DOGC núm. 4915 - 26/6/2007)

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.[12] 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 in Barcelona apply to an affiliated middle school and 88% to the closest school.

While lower and upper secondary education are two separate stages, usually they are taught in the same school ("Instituto de Educación Secundaria"). The principal is the same and classes take place in the same premises. Typically teachers are different, although some of them might change stage over time. All public schools offer the two main tracks of upper secondary education, Science and Humanities, and enrollment in guaranteed to all children who complete lower secondary education in the school. A third track, Arts, is offered in a small number of schools and there aren't guaranteed seats. In Barcelona, about 4% of the students choose Arts, 43% choose Humanities, and the remaining 53% enroll in Science. 95% of students in Science and Humanities stay in the same school for upper secondary education.

2.2 Data sources and sample creation

The *Departament d'Ensenyament* (regional ministry of education in Catalonia) provided enrollment records for schools in Barcelona, from primary education to upper secondary education, from the school year 2009/2010 to 2015/2016. In this paper I focus on public

[12]Within school, students are allocated to classes. Schools in Barcelona have from 1 to 5 classes per grade, depending on the school size. From first to third grade the curriculum is identical for all classes, although they may be taught by different teachers. In fourth grade core subjects such as Mathematics, Spanish, Catalan, and English are attended together by all students in the class; students can also chose to attend a limited number of elective subjects, whose evaluations are not part of this study. Further details on allocation of students to classes are provided in Appendix D.

middle schools.[13] For each year, data include school and class attended by the students and information on promotion or retention. Only for public schools, data contain information on final evaluations assigned by teachers at the end of the year, and 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.[14]

Moreover, I observe applications to middle schools in 2009 and 2010. For each applicant, I observe the application list, priority points for the top choice, and whether he or she enrolled in the preferred school.

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.[15] 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.[16] 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, namely about 80% of the students who enroll in a public middle school in the period under analysis.[17]

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 information can be retrieved from both sources, I impute the highest level of education,

[13] 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. 47 middle schools have sufficient data from year 2009. See Appendix B for more details.

[14] 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 about 2.5% of the total population of students who enrol for the first time in middle school.

[15] 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)

[16] 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.

[17] 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.

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). Finally, I use data from the national Tax Agency (*Agencia Tributaria*) for the average income at the postal code level.[18]

2.3 Descriptive statistics

The sample used for the analysis includes 5140 students, who begin lower secondary education in September 2009 or in September 2010, in one of 47 public middle schools in Barcelona. About 17% of them do not graduate. Most of them leave school voluntarily before completing basic education: 8.6% dropout as soon as possible, while the other stay for one or more additional years after reaching the legal age to dropout. 65.6% of the initial pool of students eventually enroll in high school (78% 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 (-0.38 standard deviations versus 0.59 for students with high parental background) and are less likely to pursue further studies. Only 43% of the initial pool of students with low educated parents enroll in high school, while 86% of students with highly educated parents do so.[19]

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 relevant differences in their attainment: 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 (71% of girls).

93% of students attend the middle school at the top of their application list. Immigrants are the subgroup with the largest share of students who did not enroll in their first choice (10%), but in all categories the overwhelming majority of pupils attend the school at the top of their list.

Disadvantaged students are more likely to have classmates with similar background and to live in a neighborhood with lower socioeconomic status index. The index of peer

[18] Those data are publicly available at https://www.agenciatributaria.es. I use data for year 2013, the first one for which information are available.

[19] I define three categories of parental background, based on parental education: "Low" if both parents have at most lower secondary education, "High" if one has tertiary education and the other has at least upper secondary education, "Average" in the remaining cases. If mother or father information are missing, I use the education level of the other parent.

quality is about 0.5 s.d. lower for students with low educated parents, and the average neighborhood SES is about 0.4 s.d. lower. Conversely, boys and girls have similar peers and live in similar neighborhoods.[20]

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% of the sample), 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 high parental education is 30 p.p; in the school at the 90th percentile 81% of students enroll in high school, while only 45% do so in the school at the 10th percentile. On the other hand, in all schools only a small share of students applied somewhere else as first choice (even in the school at the 25 percentile 90% of students had it as top choice).

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

[20] I build an index of "peer quality" based on characteristics highly correlated with performance, such as evaluations at the end of primary school and parental education. In this paper "neighborhood" is defined as a postal code zone. To construct the neighborhood SES index, I use the mean gross income in the zone, and the shares of females and males with college education. More details are given in Appendix B.

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 ability, and therefore their cognitive skills. 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 expected utility. Their flow utility at each time depends on their beliefs on cognitive skills, on their individual characteristics, on the school environment, and on their retention history.

3.2 Time line

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

- At the end of primary education, students undertake 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.[21]

- At time $t = 1$ students finish their first period in school; they receive internal grades and the notification 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.

[21]Given that students do not take any decision in the first years of lower secondary education, the model collapses them in one level. In the estimation, students are retained in level I if they are retained in first, second or third grade.

- 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. Student i starts 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 of student i in period τ at time t

The contemporaneous inputs z'_{it} for student i at time t are the unknown ability h_i, and observed individual and school variables. More specifically, individual variables include a vector of time-invariant individual characteristics $x_i = (x_{i1}, ..., x_{iJ})'$ (e.g. gender, nationality, parents' education, neighborhood quality) and a dummy rep_{it} which takes value 1 if the level is repeated for the second time. Moreover, the school environment is modeled through a vector of peer variables $p_{it} = (p_{it1}, ..., p_{itK})'$ (e.g. share of female) and a school input $\mathcal{A}_i$. 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 are retained or dropout.[22] The school input $\mathcal{A}_i$ depends only on the school in which the student i is enrolled. To simplify the notation, in the following sections I often omit the index i and use only calligraphic letters to denote school inputs.

$$C_{i,0} = z'_0\beta_0 + h_i \tag{1}$$

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

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

The notation $z'_{it}\beta_\tau$ is used for simplicity, from time $t = 1$ the functional form includes

[22] Defining peers at the class level rather than a school level appears preferable given the data used to estimate the model. In fact, in Spanish system students in the same class are exposed to the same teachers and the same contents, spending all the school time together. Moreover, this allows p_{it} to vary both over time and within school; given the limited number of cohorts I am analyzing this is a desirable feature. Class allocation is further discussed in Appendix D.

interaction between school inputs and individual time-invariant characteristics. More specifically, for $t \geq 1$:

$$z'_{it}\beta_\tau = \sum_{j=1}^{J} \beta_{\tau,x_j} x_{ij} + \mathcal{A}_i + \sum_{j=1}^{J} \beta_{\tau,sx_j}(\mathcal{A}_i x_{ij}) + \beta_{\tau,rep} rep_{it} + \sum_{k=1}^{K} \beta_{\tau,p_k} p_{itk} \tag{4}$$

This specification allows schools to affect differently students with differing characteristics (e.g. different family background) while remaining parsimonious on the the number of parameters to estimate.[23]

The cognitive ability h follows normal distribution $\mathcal{N}(0, \sigma)$ and it is uncorrelated with z_{it}. 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:

$$r_{\tau,it} = o_\tau + \lambda_\tau C_{i,\tau t} + \epsilon_{r_\tau,it}, \qquad \epsilon_{r_\tau,it} \sim \mathcal{N}(0, \rho_{r_\tau}). \tag{5}$$

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

$$r_{0,i} = C_{i,0} + \epsilon_{r_0,i}, \qquad \epsilon_{r_0,i} \sim \mathcal{N}(0, \rho_{r_0}), \tag{6}$$

and those who stay in school long enough also receive

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + \lambda_{\mathrm{II}} C_{i,\mathrm{II}t} + \epsilon_{r_{\mathrm{II}},it}, \qquad \epsilon_{r_{\mathrm{II}},it} \sim \mathcal{N}(0, \rho_{r_{\mathrm{II}}}), \tag{7}$$

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 have different grading policies: they may desing and administer more or less difficult tests and be more or less lenient

[23] A specification that estimates different school effects for students with differing characteristics would be too demanding in the current setting.

when grading. 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. Finally, similarly to nation-wide test scores, there is an exogenous normally distributed error:

$$g_{\tau,it} = \nu_\tau + \mu_\tau C_{i,\tau t} + z'_{it}\gamma_\tau + \epsilon_{g_\tau,it}, \qquad \epsilon_{g_\tau,it} \sim \mathcal{N}(0, \rho_{g_\tau}). \tag{8}$$

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. In particular, differences across schools in grading policies are captured by school inputs $\mathcal{I}$.

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. $\gamma_{\mathrm{I}} = \gamma_{\mathrm{II}} = \gamma$.[24]

I redefine evaluations as follow:[25]

$$r_{0,i} = z'_{i0}\beta_0 + h_i + \epsilon_{r_0,i} \tag{9}$$

$$g_{\mathrm{I},it} = \nu_{\mathrm{I}} + z'_{it}(\beta_{\mathrm{I}} + \gamma) + \kappa_{\mathrm{I}} I_{0,i} + \mu_{\mathrm{I}} h_i + \epsilon_{g_{\mathrm{I}},it} \tag{10}$$

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + z'_{it}\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i} + \lambda_{\mathrm{II}} h_i + \epsilon_{r_{\mathrm{II}},it} \tag{11}$$

$$g_{\mathrm{II},it} = \nu_{\mathrm{II}} + \mu(z'_{it}\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i} + \lambda_{\mathrm{II}} h_i) + z'_{it}\gamma + \epsilon_{g_{\mathrm{II}},it}, \tag{12}$$

where $I_{\tau-1,i}$ is the portion of previous cognitive skills that comes from time-varying observed covariates.[26] The coefficients in β_τ capture the cumulative effects of time invariant regressors, and the innovation of time-varying regressors. Moreover, $\mu_{\mathrm{II}} = \mu\lambda_{\mathrm{II}}$.

As a matter of notation, in the remaining of this paper I use $g_{\tau,it}$ ($r_{\tau,it}$) for internal (external) evaluation at time t in period τ. I denote g_{it} (r_{it}) the evaluation at time t, abstracting from the level, and I denote $g_{\tau,i}$ ($r_{\tau,i}$) the last evaluation in period τ, abstracting from the time.

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

[25] For convenience, I use the same Greek letters than in Section 3.3. In practice each coefficient here is a function of coefficients in 3.3.

[26] 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,

$$s(r_{\tau,it}) = h_i + \frac{1}{\lambda_\tau}\epsilon_{r_\tau,it} \tag{13}$$

$$s(g_{\tau,it}) = h_i + \frac{1}{\mu_\tau}\epsilon_{g_\tau,it}. \tag{14}$$

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.[27]

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 retention and graduation events are treated as probabilistic. At time $t = 1$ everyone is at risk of retention: students either fail and can repeat level I or are promoted to level II. At time $t = 2$, students who attended level II for a year either fail (and can repeat level II) or graduate. Similarly, at time $t = 3$, students who repeated a level in the past either fail (and leave education) or graduate.

I assume that the conditional probability takes a logit form:

[27] See DeGroot (1970)

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

Students who already repeated level I are promoted for sure to the next level. For ease of notation, I extend the definition to encompass this case: $\Pr(\text{fail}_{\mathrm{I},i2}) \equiv 0$. Moreover, I will sometime refer to the "probability of graduation" in level II, defined as $\Pr(\text{grad}_{it}) = 1 - \Pr(\text{fail}_{\mathrm{II},it})$.

The set w_{it} includes beliefs $\mathrm{E}_{i,t-1}(C_\tau)$, individual characteristics, peer characteristics and the vector of school inputs $\mathcal{I}$. 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. I assume that their leniency in retention is proportional to the leniency in grading. Students are assumed to know the parameters and form expectations over their probability of graduation using (15).

While I allow prior beliefs to enter the retention probability, I do not allow $C_{i,\tau}$ or equivalently h_i itself to enter the equation. 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. Moreover, it would make the model very difficult to treat, because students could learn about their ability through the realization of the event.[28]

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 inputs). The coefficients of the covariates capture all the motivational and non-cognitive factors which matter for the choice on top of the (perceived) level of cognitive skills. Moreover, the specification allows cognitive skills to have different effects on the flow utility of students with differing retention history.

The flow payoff of a period in lower secondary education for individual i attending level τ at time t is

$$\begin{aligned} U_{it}^M &= \phi_{M,r}\mathrm{E}_{i,t}(C_\tau) + y_{it}'\theta_M + \varepsilon_{it} = \\ &= \phi_{M,r}\mathrm{E}_{i,t}(C_\tau) + \text{ret}_{M,it}'\theta_{M,r} + x_i'\theta_{M,x} + p_{it}'\theta_{M,p} + \mathcal{T}_M + \varepsilon_{it}, \end{aligned} \tag{16}$$

[28]In fact $\text{fail}_{\tau,it}$ 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.

where $\mathcal{T}_M$ is the input on utility of the middle school in which student i is enrolled. The vector $\text{ret}'_{it} = (\mathit{stI2}_{it}, \mathit{stII3}_{it}, \mathit{ftII3}_{it})$ include three mutually exclusive dummies to capture all the possible combinations of time and repetitions. The baseline category is for students promoted to level II at time $t = 1$. *stI2* takes value 1 if the student failed at $t = 1$ and has to repeat level I at $t = 2$. *stII3* is 1 if the student has to repeat level II at $t = 3$. *ftII3* is 1 for a student who repeated the first level at $t = 2$ and can undertake second level for the first time at $t = 3$. The value of the coefficient $\phi_{M,r}$ depends on the history of retention.[29]

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

$$
\begin{aligned}
U^A_{it} &= \phi_{A,r}\mathrm{E}_{i,t}(C_{\mathrm{II}}) + y'_{it}\theta_A + \varepsilon_{it} \\
&= \phi_{A,r}\mathrm{E}_{i,t}(C_{\mathrm{II}}) + \text{ret}'_{A,it}\theta_{A,r} + x'_i\theta_{A,x} + p'_{it}\theta_{A,p} + \mathcal{T}_A + \varepsilon_{i,t},
\end{aligned}
\tag{17}
$$

where $\text{ret}'_{A,it} = (\mathit{rII}_{it}, \mathit{rI}_{it})$, with $\mathit{rII} = 1$ if the student repeated the second level, and $\mathit{rI} = 1$ if the student repeated the first level. As above, the value of $\phi_{A,r}$ depends on the retention status.

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 yields the highest expected value.

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

At the end of lower secondary education. For those who graduated, the utility of pursuing further education is simply the flow utility in (17), with $t = 2$ if retention never took place or $t = 3$ if the student was retained in either first or second level. Therefore:

$$
u_{it}|(\text{grad}_{it} = 1) = \max\Big\{0\,,\ U^A_{it}\Big\} = \max\Big\{0\,,\ v^A_{it} + \varepsilon_{it}\Big\}, \tag{18}
$$

where v^A_{it} is the utility just before observing the realization of the random shock to preferences ε_{it}.

[29] In practice the specification includes interaction terms between $\mathrm{E}_{i,t}(C_\tau)$ and the dummies ret_{it}. Therefore $\phi_{M,r}$ can take 4 different values.

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.

$$\begin{aligned} u_{i2}|(\text{grad}_{i2} = 0) &= \max\Big\{0\,,\, U^M_{i2} + \delta\,\Pr(\text{grad}_{i3} = 1)\text{E}_{i,2}(u_{i3}|\text{grad}_{i3} = 1)\Big\} = \\ &= \max\Big\{0\,,\, v^M_{i2} + \varepsilon_{i2}\Big\}. \end{aligned} \tag{19}$$

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 have to repeat level I anticipate that they will surely access level II in two periods if they stay, and then graduate with some probability:

$$\begin{aligned} u_{i1}|(\text{fail}_{I,i1} = 0) = \max\Big\{0\,,\, & U^M_{i1}|(\text{fail}_{I,i1} = 0) + \\ & + \delta\Big(\Pr(\text{grad}_{i2} = 1)\text{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 1)) + (1 - \Pr(\text{grad}_{i2} = 1))\text{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 0)\Big)\Big\} \\ & = \max\Big\{0\,,\, v^M_{i1}|(\text{fail}_{I,i1} = 0) + \varepsilon_{i1}\Big\}, \end{aligned} \tag{20}$$

$$\begin{aligned} u_{i1}|(\text{fail}_{I,i1} = 1) &= \max\Big\{0\,,\, U^M_{i1}|(\text{fail}_{I,i1} = 1) + \delta\text{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 0)\Big\} = \\ &= \max\Big\{0\,,\, v^M_{i1}|(\text{fail}_{I,i1} = 1) + \varepsilon_{i1}\Big\}. \end{aligned} \tag{21}$$

3.8 The role of schools

As detailed in previous subsections, the school environment affects students' skills, progression and choices through peer characteristics and school inputs. More specifically, I model and estimate the school input $\mathcal{A}$ on cognitive skills, the school inputs $\mathcal{T}_M$ and $\mathcal{T}_A$ on the choice of staying in middle school and on the choice of enrolling in high school after graduation, and the school leniency $\mathcal{I}$ (which affect both the difference between internal and external evaluations and the probability of retention).[30] These school inputs can be interpreted as school values added that capture the effect on outcomes of being in a given school above and beyond the quality of peers in that school. For instance, schools with

[30] In principle it is possible to separately identify $\mathcal{A}_{\text{I}}$ and $\mathcal{A}_{\text{II}}$, the school inputs on cognitive skills in first and second level. I also estimate a model that allow for different school inputs in the two levels, but their correlation is close to 1. Similarly, one could separately identify school inputs $\mathcal{I}$ and $\mathcal{F}$ on inflation of internal evaluations and probability of retention, respectively. In practice, I also estimated them separately, but again the (negative) correlation is extremely high. Therefore, I believe that the set of school inputs described in the text is the most salient.

high teacher quality have high $\mathcal{A}$, while inclusive schools that provide students at risk with additional support have high $\mathcal{T}_M$.

3.9 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.[31]

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 beliefs. 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 at time $t \in \{2, 3\}$, based on equation (11) in Subsection 3.4.3. Let assume for now that posterior belief $\mathrm{E}_i(h)$ has been computed for each student.

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + z'_{it}\beta_{\mathrm{II}} + z'_{\mathrm{I},i}\widetilde{\kappa}_{\mathrm{II}} + \lambda_{\mathrm{II}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{r_{\mathrm{II}},it}, \tag{22}$$

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. Moreover, $\widetilde{\epsilon}_{r_{\mathrm{II}},it} = \lambda_{\mathrm{II}}(h_i - \mathrm{E}_i(h)) + \epsilon_{r_{\mathrm{II}},it}$. Under the assumption that educational choices depend only on the belief about ability, the errors $\widetilde{\epsilon}_{r_{\mathrm{II}},it}$ is uncorrelated with the 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 12):

$$g_{\mathrm{II},it} = \nu_{\mathrm{II}} + z'_{it}(\mu\beta_{\mathrm{II}} + \gamma) + z'_{\mathrm{I},i}\mu\widetilde{\kappa}_{\mathrm{II}} + \mu_{\mathrm{II}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{g_{\mathrm{II}},it}, \tag{23}$$

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 ap-

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

plication of ordinary least square to

$$g_{\mathrm{I},it} = \nu_{\mathrm{I}} + z_{it}'(\beta_{\mathrm{I}} + \gamma) + z_{0,i}'\widetilde{\kappa}_{\mathrm{I}} + \mu_{\mathrm{I}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{g_{\mathrm{I}},it}, \tag{24}$$

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

$$r_{0,i} = z_{i0}'\beta_0 + h_i + \widetilde{\epsilon}_{r_0,i} \tag{25}$$

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}.

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 $(\rho_{r_0}, \rho_{g_{\mathrm{I}}}, \rho_{g_{\mathrm{II}}}, \rho_{r_{\mathrm{II}}})$. Those are identified from the history of signals, particularly the covariance of evaluations within and over time. In particular, σ can be inferred from the covariance of the residuals of $g_{\mathrm{II},it}$ and $r_{\mathrm{II},it}$ on the observable covariates. 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.[32]

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, $\text{fail}_i = (\text{fail}_{it})_t$ the vector of retention/graduation events, and $o_i = (o_{it})_t$ the vector of evaluations observed by i.[33] The student takes $T_d \in \{1, 2, 3\}$ decisions, receives $T_{\mathrm{f}} \in \{1, 2, 3\}$ 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, namely $d_i = (1, 0)$, she is notified retention at $t = 1$ and she would be promoted at $t = 2$ if she stayed, i.e. $\text{fail}_i = (0, 1)$, and she observes evaluations $o_i = (r_{0,i}, g_{\mathrm{I},i1}, g_{\mathrm{I},i2})$.

[32]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, $\mathrm{Var}(r_{\mathrm{II},it}) = \lambda_{\mathrm{II}}^2\sigma + \rho_{\mathrm{II}}^r$.

[33]o_{it} is a vector containing one evaluation at $t = 0$ and in level I, and up to two evaluations in level II. Recall that I use $g_{\tau,it}$ ($r_{\tau,it}$) for internal (external) evaluation at time t in level τ, while I denote g_{it} (r_{it}) the evaluation at time t, abstracting from the level, and I denote $g_{\tau,i}$ ($r_{\tau,i}$) the last evaluation in level τ, abstracting from the time.

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

$$L_i = L(d_i, \text{fail}_i, o_i) = L(d_{i1}, ..., d_{iT_d}, \text{fail}_{i1}, ..., \text{fail}_{iT_\text{f}}, g_{i1}, ..., g_{iT_d}, r_{i0}, ..., r_{iT_r}) \tag{26}$$

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

$$\begin{aligned}L_i = &\int L(r_{i0}|h)L(\text{fail}_{i1}|h, r_{i0})L(g_{i1}|h, r_{i0})L(d_{i1}|h, r_{i0}, g_{i1})...L(d_{iT_d}|h, o_i, d_{i1}, ..., d_{it_d-1})\, \boldsymbol{\phi}(h)dh = \\ &\Big(L(d_{i1}|r_{i0}, g_{i1})...L(d_{iT_d}|o_i, d_{i1}, ..., d_{iT_d-1})\Big) \times \Big(L(\text{fail}_{i1}|r_{i0})...L(\text{fail}_{iT_\text{f}}|o_i, d_{i1}, ..., d_{iT_d-1})\Big)\times \\ &\times \int L(o_{iT_d}|h, d_{i1}, ..., r_{i0}, ...)...L(r_{i0}|h)\, \boldsymbol{\phi}(h)dh\end{aligned} \tag{27}$$

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:

$$\log L_i = \log L_{i,d} + \log L_{i,\text{fail}} + \log L_{i,o} \tag{28}$$

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.

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 a logit model for the probabilities of failure, and a model of dynamic choices with logistic errors. More details are provided in subsections 4.3 and 4.4.

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 evaluations and the parameters.

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

$$\arg\max_{\zeta,\sigma} \sum_i \log L(o_i; \zeta, \sigma) = \arg\max_{\zeta,\sigma} \sum_i \log \int L(o_i; \zeta, \sigma | h) \boldsymbol{\phi}(h)\, dh \tag{29}$$

The main point behind this application of the EM algorithm is that if $\widehat{\zeta}$ is a maximizer for (29), then it also solves

$$\arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma | h) \boldsymbol{\psi}_i(h)\, dh \tag{30}$$

Therefore for a given value of σ, $\widehat{\zeta}$ can be retrieved using (30) rather than (29). $\widehat{(\zeta, \sigma)}$ can be estimated using an iterative algorithm: at each iteration k, first (E-step) posterior distributions $\boldsymbol{\psi}_i^k(h)$ are computed for all individuals using previous iteration estimates ζ^{k-1}. Then (M-step) estimates of pararameters ζ^k are computed as solution of (30).

Appendix (C) provides a more detailed theoretical motivation. Next paragraphs describe the estimation procedure.

4.2.1 E-step

At step k, posterior distribution $\boldsymbol{\psi}_i^k(h)$ is 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 ω_i^k the posterior variance. Moreover, at the end of E-step, the estimate for the population variance is updated; this new σ^k is used at the beginning of next step $k+1$. The updating formula for σ^k is retrieved using the law of total variance:[34]

$$\sigma = \mathrm{E}\left(\omega_i + \mathrm{E}_i(h)\mathrm{E}_i(h)'\right), \tag{31}$$

The sample equivalent at step k is computed as

$$\widehat{\sigma}^k = \frac{1}{N}\left(\omega_i^k + \mathrm{E}_i^k(h)^2\right) \tag{32}$$

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

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

4.2.2 M-step

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

$$\arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma^{k-1}|h) \boldsymbol{\psi}_i^k(h)\, dh \tag{33}$$

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

$$\sum_i \int \log L(o_i; \zeta, \sigma^{k-1}|h) \boldsymbol{\psi}_i^k(h)\, dh = \tag{34}$$

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

$$= \sum_i \int \log L(r_{0,i}; \zeta, \sigma^{k-1}|h) \boldsymbol{\psi}_i^k(h)\, dh + \sum_{it} \int \log L(g_{I,it}; \zeta, \sigma^{k-1}|h) \boldsymbol{\psi}_i^k(h)\, dh + \tag{35}$$

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

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:[35]

[35] It is easy to see how to derive (36) from (35). For instance the contribution of the first region-wide test is given by:

$$\int \log L(r_{0,i}; \zeta, \sigma^{k-1}|\eta) \psi_i^k(h)\, dh = \int \log \left(\frac{1}{\sqrt{2\pi\rho_{r_0}}} \exp\left(-\frac{(r_{0,i} - h - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0})^2}{2\rho_{r_0}} \right) \right) \psi_i^k(h)\, dh =$$

$$= \int \left(-\frac{1}{2}\log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}} (r_{0,i} - h - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0})^2 \right) \psi_i^k(h)\, dh =$$

$$= -\frac{1}{2}\log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}} \mathrm{E}_i^k \left((r_{0,i} - h - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0})^2 \right) =$$

$$= -\frac{1}{2}\log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}} \left(\mathrm{Var}_i^k \left(r_{0,i} - h - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0} \right) + \left(\mathrm{E}_i^k (r_{0,i} - h - x_i'\beta_{x,0} - p_i'\beta_{p,0}) \right)^2 \right) =$$

$$= -\frac{1}{2}\log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}} \left(\omega_i^k + \left(r_{0,i} - \mathrm{E}_i^k(h) - x_i'\beta_{x,0} - p_{i,0}'\beta_{p,0} \right)^2 \right)$$

$$
\begin{aligned}
-\sum_i logL_i \;=\; & \sum_i \frac{1}{2}\log(2\pi\rho_{r_0}) + \frac{1}{2\rho_{r_0}}\left(\omega_i^k + \left(r_{0,i} - (\mathrm{E}_i^k(h) + z_{i0}'\beta_0)\right)^2\right) + \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{g_{\mathrm{I}}}) + \frac{1}{2\rho_{g_{\mathrm{I}}}}\left(\omega_i^k + \left(g_{\mathrm{I},it} - \left(\nu_1 + \mu_{\mathrm{I}}\mathrm{E}_i^k(h) + z_{it}'(\beta_{\mathrm{I}} + \gamma) + \kappa_{\mathrm{I}} I_{0,i}\right)\right)^2\right) + \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{r_{\mathrm{II}}}) + \frac{1}{2\rho_{r_{\mathrm{II}}}}\left(\omega_i^k + \left(r_{\mathrm{II},it} - \left(o_{\mathrm{II}} + \lambda_{\mathrm{II}}\mathrm{E}_i^k(h) + z_{it}'\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i}\right)\right)^2\right) + \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{g_{\mathrm{II}}}) + \frac{1}{2\rho_{g_{\mathrm{II}}}}\left(\omega_i^k + \left(g_{\mathrm{II},it} - \left(\nu_{\mathrm{II}} + \mu\lambda_{\mathrm{II}}\mathrm{E}_i^k(h) + \mu(z_{it}'\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i}) + z_{it}'\gamma\right)\right)^2\right)
\end{aligned} \tag{36}
$$

The total likelihood in (36) 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 (36) 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_{i0}. This provides us with updated estimates β_0^k, and allows the computation of $I_{i0}^k = s_{i0}\beta_{s,0}^k$, which is used in next step.
    - Update variances $\rho_{r,0}^k$, using the sample equivalent of $\mathrm{E}_\epsilon(\mathrm{E}_h(\epsilon_{r_0,i}|r_{0,i}))$:

    $$\mathrm{Var}(\epsilon_{r_0,i}) = \mathrm{E}(\epsilon_{r_0,i}^2) = \mathrm{E}(\mathrm{E}((r_{0,i} - h_i - z_{i0}'\beta_0|r_{0,i})^2)) = \tag{37}$$

    $$= \mathrm{E}\left(\int (r_{0,i} - h - z_{i0}'\beta_0)^2\, \psi_i(h)\, dh\right) = \tag{38}$$

    $$= \mathrm{E}\left[\mathrm{Var}_i(h) + \left((r_{0,i} - \mathrm{E}_i(h) - z_{i0}'\beta_0)^2\right)\right] \tag{39}$$

    which, in each iteration k, can be estimated from the sample as

    $$\rho_{r,0}^k = \frac{\sum_i \left(\omega_i^k + (r_{0,i} - E_i^k(h) - z_{i0}'\beta_0^k)^2\right)}{N} \tag{40}$$

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

I estimate the parameters of the logit model described in Section 3.5. Regressors include beliefs on cognitive skills and school leniency effects, which are estimated using results from the previous stage. More specifically, I use the estimated parameters $\widehat{\zeta}$ and individual posterior distributions of ability $\widehat{\psi}_i$ to compute beliefs about cognitive skills at each point in time. Moreover, the estimated school effects $\widehat{\mathcal{I}}$ are used as a measure of school leniency.

The estimated coefficients can then be used to compute the probability of repeating level I at time $t = 1$ and the probability of graduation in the following time periods. The individual probability of failure in level II enters the student's maximization problem, while the probability of retention in level I does not, given that it happens before any decision has to be taken. However the estimation of the latter is necessary for simulations and conterfactual analyses.

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.

By assumption, 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

$$\begin{aligned}\mathrm{E}_\varepsilon(u_{it}|v_{it}^A) = \log\big(\exp(v_{it}^A)+1\big) = \\ = \log\left(\exp(\phi_{A,r}\mathrm{E}_{i,t}(C_{\mathrm{II}}) + y_{it}'\theta_A) + 1\right)\end{aligned} \tag{41}$$

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:

$$\mathrm{E}_{i,t}(h) = (\widehat{h}_{it-1}, \mathbf{s}_t')\mathbf{c}_t = c_{0,t}\widehat{h}_{it-1} + c_{1,t}s_{r,t} + c_{2,t}s_{g,t} \tag{42}$$

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

$$\begin{aligned}\mathrm{E}_{i,t-1}(u_{it}|\mathrm{grad}_{it}=1) = \int \log\left(\exp(v_{i,A})+1\right)\cdot \widehat{\psi}_{it-1}(s_{g,t}, s_{r,t})\, d\mathbf{s}_t = \\ = \int \log\Big[\exp\left(\phi_{A,r}\left(k_{\mathrm{II}}I_{t-1} + z_{it}'\beta_{\mathrm{II}}\right) + y_{it}'\theta_A\right)\cdot \exp\Big(\phi_{A,r}c_{0,t}\widehat{h}_{it-1}\Big)\cdot \\ \cdot \exp\left(\phi_{A,r}(c_{1,t}s_{r,t} + c_{2,t}s_{g,t})\right) + 1\Big]\cdot \widehat{\psi}_{it-1}(s_{g,t}, s_{r,t})\, d\mathbf{s}_t\end{aligned} \tag{43}$$

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:

$$v_{i2}^M(\widehat{h}_{i2}, z_{i2}) + \varepsilon_{i2} = U_{i2}^M + \delta\widehat{\Pr}(\mathrm{grad}_{i3}=1)\int \log\big(\exp(v_{i3}^A)+1\big)\cdot \widehat{\psi}_{i2}(s_{g,3}, s_{r,3})\, d\mathbf{s}_3 \tag{44}$$

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:

$$v_{i,1}^M|(\mathrm{fail}=1) + \varepsilon_{i,1} = U_{i,1}^M + \delta E_{i,1}(u_{i2}|\mathrm{grad}_{i2}=0) \tag{45}$$

$$\begin{aligned}v_{i,1}^M|(\mathrm{fail}=0) + \varepsilon_{i,1} = U_{i,1}^M + \delta\Big(\Pr(\mathrm{grad}_{i2}=1)\mathrm{E}_{i,1}(u_{i2}|\mathrm{grad}_{i2}=1)) + \\ + (1-\Pr(\mathrm{grad}_{i2}=1))\mathrm{E}_{i,1}(u_{i2}|\mathrm{grad}_{i2}=0)\Big)\end{aligned} \tag{46}$$

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_{i2}^M, 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:

$$p_{it}(d_{it} = 1|d_{it-1} = 1) = \frac{\exp(v_{it})}{1 + \exp(v_{it})} \tag{47}$$

I maximize the total loglikelihood to estimate the parameters, following Rust (1987).[36]

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 I allow for the two different scenarios in the computation of expected utility.

4.6 Standard errors

Standard errors are estimated using a bootstrap procedure with 200 replications. Let N_s be the number of students in the sample enrolled in school s. In each replication, for each school s, N_s individuals are sampled with replacements.[37] An alternative approach does not impose that the number of students in the school stay constant across iterations, and simply draw with replacement N students from the total sample at each iteration. I estimated standard errors using also this alternative approach and results are quite similar, in particular the significance of the coefficients never change.

5 Results

In this section I discuss the results of the estimation of the model presented in Section 3. The main findings are the following:

- *Cognitive skills.* About half of the total variance of cognitive skills is due to unobserved ability, however evaluations are informative and posterior individual variance

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

[37]For two schools, in very few cases the random sample did not contain any dropout. In this case, I re-sample again the observations for the school. This happen in less than 4% of the iterations for one school and in less than 0.4% for the other school.

shrink rapidly.[38] Parental education is the most important observed determinant of cognitive skills accumulation: having both parents with tertiary degree rather than basic education is associated with an improvement in skills by about 1 s.d. School environment is quite relevant as well: being with higher quality peers increases performance, 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 almost 0.4 s.d.). School effects are quite homogeneous across individual characteristics.

- *Educational choices.* Cognitive skills (or, more precisely, beliefs about them) are the most important determinant of the choices. The school environment has a large direct impact, for instance being in a school at the 75 percentile of the distribution of school inputs rather than at the 25 affects the choice of pursuing further education almost as much as having 0.5 s.d. higher cognitive skills. Peer quality is associated with a modest decrease in the flow utility from staying in middle school, perhaps due to ranking concerns.

- *Retention.* School environment affects probability of retention and graduation above and beyond its impact on cognitive skills. For instance, being in a school at the 75 percentile of the distribution of school leniency rather than at the 25 decreases probability of retention as much as an increase of cognitive skills by 0.5 s.d. When choosing whether to pursue further education, retained students exhibit a flatter flow utility in cognitive skills, therefore there is a larger gap among retained and non-retained students with higher cognitive skills than with lower cognitive skills. For the choice of academic upper secondary education, the value of the flow utility of retained students relatively to their outside option is significantly lower than the value of non-retained students at any given level of cognitive skills.

In Subsections 5.1 to 5.3 I discuss the estimated parameters; in Subsection 5.4 I discuss the fit of the model.

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 0.28. 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 the unknown ability h is uncorrelated with the regressors, it is possible to decompose it in variance due to observable characteristics, and variance due to unknown ability.

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

Second and third columns 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 h; more specifically, about 49% before starting middle school and in the first level of middle school, and about 58% in the second level.

For comparison, the fourth and fifth columns of the table contain the variance of beliefs about unknown ability and about total cognitive skills in each level. In levels I and II variances of beliefs are similar to the previous estimates, although slightly smaller. This is coherent with the finding in Panel B, which contains posterior variance in each period for every possible set of signals. Evaluations are quite informative, in fact the posterior variance is about 0.17 at time 0, it shrinks to 0.06 at time $t = 1$, and it is further reduced to less than 0.04 at $t = 2$. It is 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. Summarizing, when students take the first choice at time $t = 1$, they generally have pretty accurate beliefs about their ability and therefore about their cognitive skills. This suggests that their decisions would be unlikely to change if they observed their exact skills level rather than relying on their beliefs. I confirm this intuition in Section 6 when I compare their outcomes with a counterfactual scenario without uncertainty.

Table 4 presents estimates for the parameters governing evaluations, and therefore cognitive skills, before starting middle school (β_0) and in level I and II (β_{I} and β_{II} respectively). The last column of the table contains the parameters that capture differences between internal and external evaluations (γ). Since the evaluations have been standardized, the coefficients of the individual characteristics can be interpreted as standard deviation change.

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, in level II having a mother with tertiary rather than primary education increases performances by more than 0.6 s.d., having a highly educated father increases performances by more than 0.3 s.d. This means that a student with highly educated parents is expected to receive evaluations about 1 s.d. higher than a classmate with identical characteristics and unobserved ability, but whose parents only have primary education. Students whose parents completed high school (i.e. "average" parental background) are in an intermediate position: everything else equal they perform about 0.5 s.d. better than those with low educated parents.

Immigrant students do somewhat worse than natives. Boys and girls do similarly. While, everything else equal, there aren't relevant differences in the cognitive skills accumulation of boys and girls, it is interesting to remark that being a female is associated with a large premium for internal evaluations (almost 0.4 s.d.). Being younger when entering primary education is a disadvantage, although the gap is decreasing over time. In fact, 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 skills before starting middle school and with a increase by 0.14 s.d. at the end of it.[39] Starting middle school with a one-year delay is associated with lower performance (up to -0.7 s.d. at the end of middle school).

Repeating a level for a second time has a positive effect on cognitive skills, especially in second level. In fact final external evaluations are more than 0.5 s.d. larger for retained students everything else equal (the effect is about 0.2 s.d. in first level).

The quality of the neighborhood of residence, as measured by an index that takes in account average wealth and education, has a small positive effect on performance: an increase by 1 s.d. of the index is associated with an improvement of performance by at most 0.08 s.d.[40]

The set of covariates include a dummy for students who are attending a school which was not their top choice in the application list. Coefficients are always insignificant and negligible in size. This provide some evidence that students with differing preferences for schools do not have a different development of their cognitive kills, everything else equal.

Peer quality has a positive effect on cognitive skills, while the share of female does not does not appear to have any relevant effect. In order to capture non linear peer effect, I use a polynomial of degree 3 for each peer regressor. To facilitate the interpretation of the results, Table 4 displays the effect on evaluations associated with having peers at the mean rather than 1 s.d. below the mean. Having peers of average quality rather than 1 s.d. below average is associated with an improvement by almost 0.3 s.d. in level I. Moreover Figure 2 plots the effect for each value of the peer variable. As clear from the figure, being with better peers always increases evaluations, but the effect is more pronounced in the tails of the distribution. The pattern is similar in level II, but the effect are smaller in size, although still significant. On the other hand, the positive effect is offset in internal evaluations. This is aligned with the finding in Calsamiglia and Loviglio (2019a) that in Catalonia teachers' evaluations are deflated in the presence of better quality peers, i.e. that teachers are somewhat "grading on a curve". The proportion of female peers has no effect on skills at any point of the distribution.

To summarize the estimated school inputs, Table 4 displays the difference between the school input on cognitive skills $\mathcal{A}$ at the 75th percentile and at the 25th percentile (in-

[39] Calsamiglia and Loviglio (2019b) document the disadvantage of being younger at school entry throughout compulsory education in Catalonia.

[40] The indexes of peer quality and neighborhood quality are described in Section 2.3 and Appendix B.2. I also estimated the model with a vector of variables for peer characteristics, results are aligned although less precisely estimated. Moreover, I estimated the model with separate variables for wealth and education in the neighborhood, results are fully aligned. Therefore, I opted for the most parsimonious specification.

terquantile range), the difference between the 80th and 20th percentile, and the standard deviation. There is a sizable variation across schools in their effects: the interquantile range for the school input on cognitive skills is almost 0.4 s.d. both in the first and in the second level. The interquantile range for the school input $\mathcal{I}$, which capture school grading policies, is also sizeable; moreover the school inputs $\mathcal{I}$ and $\mathcal{A}$ have large negative correlation (about -0.7). In other words, schools that improve the cognitive skills the most tend to have a stricter grading policy for internal evaluations.[41]

The empirical specification includes interactions between the vectors of school inputs $\mathcal{A}$ and $\mathcal{I}$, and the dummies for gender, nationality, and parental education, to account for differing school effects across socioeconomic groups. If, for instance, schools matter more for the cognitive skills development of girls than boys, this would be captured by the coefficient of the interaction term. To ease the interpretation of the results, Table 4 report for each individual characteristic the change due to an increase of 1 s.d. in the school effect. The estimated interaction effects are quite small and insignificant. This means that the effect of the school environment on cognitive skills development is about the same for all students enrolled in a given class, no matter their gender, nationality or parental background.

5.2 Retention and graduation

Table 5 contains estimates of the parameters governing the logit model for retention (introduced in Section 4.3).[42] Not surprisingly, higher cognitive skills decreases the probability of retention. Being female decrease the probability of retention above and beyond the skills level. Also having a highly educated father has a negative effect, but it is smaller than the one due to gender.[43]

As shown by Figure 2, being with better peers increases the probability of retention for a given level of skills. This is consistent with the negative effect of peer quality on internal evaluations discussed in the previous section: teachers may be more prone to fail a student who is behind his peers than a student with identical skills level who is close to the median of his class.

As explained in Section 4.3, the vector of school leniency effects $\mathcal{I}$ is used as regressor. There are large differences across schools in the probability of failing for a given level of

[41] This is coherent with the the evidence in Calsamiglia and Loviglio (2019a) that in Catalonia schools in which average external evaluations are higher exhibit stricter grading policy.

[42] The table has the same structure of Table 4: for peer regressors it shows the effect of increasing the variable 1 s.d. below average to the average; for school effects it reports differences between percentiles of the distribution.

[43] 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.

cognitive skills. For instance, 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 0.4 s.d. Increasing school leniency $\mathcal{J}$ by 1 s.d. has about the same effect on retention than increasing the school inputs on skills $\mathcal{A}$ by 1 s.d. (therefore improving skills by almost 0.3 s.d.).

Results are similar if the school leniency on retention policy is estimated directly using school dummies. The correlation between the estimated vector of school effects and $\mathcal{J}$ is larger than 0.8. Given that rules for grading and for retention are most likely designed together at the school level, I prefer to estimate only one parameter for each school.

5.3 Educational choices

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.

For a given level of cognitive skills, students with below average peers are more likely to choose to stay in middle school than those with above average peers. 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.[44] Figure 2 shows that the effect is non linear: it flattens for values of peer quality above the average. No relevant peer effects are detected for the choice of enrolling in high school.

Both school inputs $\mathcal{T}_M$ on the choice of staying in middle school and school inputs $\mathcal{T}_A$ on the choice of enrolling in high school are sizeable. In both cases, being in a school with school input at the 75 percentile rather than at the 25 percentile increases the flow utility as much as improving cognitive skills by 0.5 s.d. Increasing school inputs $\mathcal{T}_M$ or $\mathcal{T}_A$ by 1 s.d. has a larger impact on flow utility than the indirect effect obtained through an increase of school input $\mathcal{A}$ on cognitive skills by 1 s.d.

It is important to stress that although school inputs have a linear effect on the flow utility, the school environment may affect differently classmates who differ in their individual characteristics or cognitive skills. 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.[45] For instance, the school environment may matter differently for the educational decisions of students with low or highly educated parents

[44] 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.

[45] 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

even if they are attending the same class. In next section, I discuss extensively how the same change in the school environment can affect differently the choices of students with differing parental background.

Retention during lower secondary education has a large negative effect on the flow utility for the choice of high school. The interactions of cognitive skills with dummies for repeating level I or level II also have negative coefficients, thus the total coefficient of cognitive skills for retained students is smaller than for students who are progressing regularly.[46] In other words, results suggest that being retained decreases the weight that individuals give to cognitive skills when making their choice. Figure 1 helps to illustrate this point. 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 increases relatively less the utility of retained students than the utility of non-retained students.[47] 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 higher education.

The effect of retention on the choice of staying in school goes in a similar direction, but coefficients are generally smaller in size and not precisely estimated.

5.3.1 School inputs on choices and school leniency

The estimate school inputs $\mathcal{T}_M$ and $\mathcal{T}_A$ allow us to quantify how schools affect students' choices above and beyond their cognitive skills. In this subsection I move one step further and isolate the school contributions $\mathcal{M}_M$ and $\mathcal{M}_A$ that are uncorrelated with the school grading policy and can be thought as purely affecting students' motivation.

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.

[46]The coefficient of the interaction with the dummy for repeating level I is significant at 1%, the other is slightly smaller and only marginally significant.

[47]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.

It is worth recalling that school leniency $\mathcal{I}$ enters students' educational decisions through the probability of retention and graduation: in less demanding schools it is easier to avoid retention and complete lower secondary education in time, therefore the expected utility is higher. $\mathcal{I}$ also affect internal evaluations, which however do not enter directly the utility of the students. In practice, one feature not explicitly discussed so far is that students may prefer to have higher rather than lower grades, for a given level of cognitive skills. This may happen for instance if family members reward good grades or when GPA counts towards University admission.[48] Then, the schools effects $\mathcal{T}_M$ and $\mathcal{T}_A$ would capture any direct effect of $\mathcal{I}$ on the educational choices.

Disentangling the contribution of school leniency $\mathcal{I}$ from other channels may be important, especially because it is unclear if grade inflation would be a desirable policy. Therefore, I decompose $\mathcal{T}_M$ and $\mathcal{T}_A$ in two orthogonal components:

$$\begin{aligned} \mathcal{T}_{\mathcal{M}} &= a_M \mathcal{I} + \mathcal{M}_M, \qquad \mathcal{M}_M \perp \mathcal{I} \\ \mathcal{T}_A &= a_A \mathcal{I} + \mathcal{M}_A, \qquad \mathcal{M}_A \perp \mathcal{I} \end{aligned} \tag{48}$$

$\mathcal{M}_M$ and $\mathcal{M}_A$ can be interpreted as the school contributions to choices unrelated to the school grading policy. I estimate (48) with OLS using as dependent variable the estimated school input on choice and as a regressor the estimated school input on the grading policy. The residual at the school level are used to estimate $\mathcal{M}_M$ and $\mathcal{M}_A$; by construction they are orthogonal to the estimated $\mathcal{I}$. The estimated coefficients a_M and a_A of $\mathcal{I}$ and the usual statistics on school inputs are shown in the last lines of Table 6.

Both $\mathcal{I}$ and the orthogonal inputs $\mathcal{M}_M$ and $\mathcal{M}_A$ explain a sizeable portion of the aggregate school inputs on choices. About 50% of the variance of $\mathcal{T}_M$ is due to $\mathcal{M}_M$, the share rises to 75% for $\mathcal{T}_A$ and $\mathcal{M}_A$.

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 1000 copies of each individual at time 0 (i.e. of her time invariant characteristics, primary school attended, and middle school in which she enrols). 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, in particular their cognitive skills and their choices.

Table 7 reports empirical frequencies ("data" columns) and frequencies predicted using

[48] In Spain high school GPA counts for about 50% of the total score that determines University admission. As explained in Section 2.1, almost all students stay in the same school for upper secondary education. Thus students may take $\mathcal{I}$ as a proxy for the leniency of the high school.

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-13 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 in 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, gender, ...). 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 almost identical to the empirical one, while the retention rate in second level is somewhat higher (about 1.5 p.p. more).

Next, I investigate 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

As shown in Section 5.1, having more educated parents is associated with a larger growth of cognitive skills. Given that they have higher skills, students with high parental background are less likely to be retained and more likely to pursue further studies.[49] Results in previous section also suggest that the school environment, i.e. peer quality and school inputs, can substantially affect cognitive skills, retention, and choices. The purpose of this section is to study how important the school environment is for the outcomes of students with differing parental background. My aim is to answer questions such as: what would gain or lose a student with low educated parents if his school environment was changed, for instance improving some of the school inputs or changing peer quality? And what about a student with highly educated parents?

Before presenting more in details the simulations that I implement, it is worth discussing how the school environments experienced by students with low or high parental background compare. On one hand, on average, students with low parental background are exposed to different peers than students with high parental background: in fact, the

[49] High parental background also slightly decrease the probability of retention for a given level of skills. Parental background has only minor direct effects on choices. The large effect of parental backround on skills is the main driver of the differences in choices and attainment between as student with highly educated parents and a student with low educated parents, everything else equal.

latter are more likely to have higher quality peers (Section 2.3).[50] On the other hand, most of the school inputs that they receive are, on average, similar. In fact, as shown in Table 8, the average grading policy $\mathcal{I}$ is almost identical for the two groups. $\mathcal{A}$ and $\mathcal{T}_M$ are slightly larger for students with low educated parents, but the differences are quite small in size and insignificant. The only exception is the school input on the choice of enrolling in upper secondary education: students with low educated parents are more likely to attend schools that encourage high school enrollment ($\mathcal{M}_A$ is significantly larger for them).

It is important to remark that all schools that I observe in the data have a diverse pool of students, although the shares of students with low or highly educated parents vary considerably across schools. In other words, the modal student with low parental background attends a school with many children like him, but there are also several low parental background students enrolled in schools where most children have a higher parental background.

I use the model and the estimated parameters to study cognitive skills development, choices, and attainment 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 performance and attainment due to the school environment (Subsection 6.2). Moreover, I explore differences in educational patterns of retained and non-retained students who have identical cognitive skills before retention (Subsection 6.3). Finally, I study students' outcomes under counterfactual scenarios in which selected school inputs are improved, or the length of compulsory education is increased (Subsection 6.4).

To make results fully comparable across parental background, in most of the analyses I keep constant all individual characteristics beside parental education. In other words, I compare the outcomes of two fictional students, *type L* and *type H*, who have identical innate ability and identical observable characteristics (e.g. same gender or nationality) with the exception of their parental background. Type L has low educated parents, more specifically both mother and father attained at most lower secondary education. Type H has highly educated parents, more specifically both mother and father completed tertiary education.[51]

[50] Results in Section 5 show that having higher quality peers is associated with an improvement in cognitive skills, but also, for a given level of skills, with and increase in the probability of retention and in some cases with an increase in the probability of dropout. Therefore it is important to assess what is the total effect of peer quality on students' choices and attainment.

[51] Overall 18% of students in the sample have both parents with at most lower secondary education. The other students who belong to the group of students with "low educated parents" are those with one parent with basic education and missing information for the other parent. 16% of students have both parents with tertiary education. The other students with "highly educated parents" are those with one parent with tertiary education and the other parent missing or with upper secondary education.

The reference individual characteristics for the simulations discussed in this section are: male, Spanish, born in the middle of the year, began lower secondary education at 12 years old, attends the middle school at the top of their application list. Primary school effect, cohort effect, and neighborhood quality are all set at the average value. I replicated all the analyses with several other set of characteristics, particularly for female and immigrant students. Overall results exhibit similar patterns, and differences go in the expected direction. For instance, female are less likely to be retained, and they choose to pursue further education more, immigrants are more likely to dropout.

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 input among students with the same parental background. The first column of the table reports also results for a student with average parental background for comparison. The first three lines of the table reports the most compelling statistics: rate of graduation, rate of enrollment in high school and cognitive skills at time 1 (i.e. computed before any dropout occurs). The following lines contain additional statistics on choices, performance and attainment.

Not surprisingly, there is already a gap in skill level of type L and type H when they start lower secondary education and it widens in middle school. Graduation rate for Type L is 77%, and the rate of enrollment in high school is less than 40%. Type L students have 31% chances of being retained in first period, 11% of dropping out immediately, and 20% of dropping out at time 2 if they are still is school but did not finish. Dropout is high despite the fact that those who choose to stay have very high probability of succeeding (95%). Among those who graduate, 51% enroll in high school. Conversely, Type H graduation rate is 99% and enrollment in high school is 93%. Only 3% of type H students are retained in first period, 0.7% dropout immediately and 4% of retained students dropout at time $t = 2$. 94% of those who graduate enroll in high school.

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

When type L attends a school with the average school inputs of type H, he acquires about the same level of cognitive skills. He is slightly more likely to dropout and the graduation rate drop by about 1.5 percentage points. The probability of attending high school drop by almost 6 p.p. Those effects are mainly driven by the lower school inputs on the choice of undertaking upper secondary education. Not only it directly affects enrollment among graduate students, but it also makes dropout more appealing because it reduces the continuation value of staying in education.[52] In column (2) graduation rate of type H is basically unaffected, while the probability of attending high school increases slightly (about 1.5 p.p.).

A change in peers composition affects positively the outcomes of type L, as shown in columns (3). The large gain in cognitive skills (from -0.46 to -0.2) more than compensate the negative effect that peer quality has on promotion and choice of staying in school; in fact graduation rate increases by 3 p.p. Enrollment in high school exhibit a sizable improvement by more than 9 p.p. While the increase in cognitive skills is reflected by the improvement in external evaluations, internal evaluations do not improve due to the "grading on the curve" effect. Type H experiences a symmetric drop in cognitive skills, and negative but smaller effects on attainment.

In columns (4), each type is given the overall average environment of the other type. The effects on cognitive skills are similar to those in columns (3), while there are smaller effects on graduation rate and enrollment in high school.

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 affect positively his choices and attainment. 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 a given student may experience a quite different educational path depending upon the specific school in which he is enrolled. I study the relevance of school environment school by school in next Subsection 6.2.

The exercise presented here and the results in Section 5 suggest that retention can be important for subsequent choices and attainment. Subsection 6.3 provide further insights.

6.2 Variation of outcomes across schools

I use the model to simulate, for each single school, the outcomes that a given type of student would have if enrolled there. This allows to quantify the importance of the school environment on achievements and choices of students with differing family background. More specifically, I simulate 10000 fictitious students for both type L and type H, and

[52] I also replicated the Table changing only value for $\mathcal{T}_A$ and results go in the same direction.

I draw shocks to evaluations, preferences, and retention events; then I compute their outcomes in every school in the sample. I use the average peer characteristics in the school for peer variables, therefore results are representative of the outcomes that a student of a given type would have if enrolled in a given school. It is important to stress 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 the school environment is for each type.

The left half of Table 10 contains results for type L, the right half contains results for type H. As in Table 9, the first three lines report graduation rate, enrollment in high school, and cognitive skills at time $t = 1$, the following ones show additional statistics. The first column for each type contains the difference between the expected outcome in the school at the 80 percentile and at the 20 percentile, the second column shows the interquantile range, and the third one the standard deviation.

For both type L and type H, cognitive skills accumulation vary widely from one school to another. A student enrolled in the school at the 75 percentile can expect to acquire almost 0.5 s.d. higher skills than in the school at the 25 percentile. Cognitive skills are almost linear in school inputs (as shown in Section 5.1), therefore it is not surprising that the school environment affects similarly the cognitive skills development of students with differing parental background.

Conversely, the school attended can make a large difference for the attainment of type L, but its impact on the attainment of type H is relatively small. For type L, graduation is 12 percentage points more likely in the school at the 75 percentile than in the school at the 25 percentile. Differences across schools are due to the differences in cognitive skills accumulation, but also to differences in retention probability (the interquantile range for retention at $t = 1$ is about 10 p.p.) and dropout rate (the interquantile range is almost 7 p.p. at $t = 1$ and almost 8 p.p. at $t = 2$). Similarly, the probability of pursuing upper academic education is much higher in some schools than in others, both unconditionally and conditionally on graduation (the interquantile range is about 12 p.p.). For type H, the interquantile range for graduation is less than 1 p.p., it is 4 p.p. for enrollment in high school (3 p.p. conditional on graduation). Given his parental background, this guy reaches an above average level of skills in each school and he is almost always promoted. Therefore his flow utility from pursuing further studies is typically very high: a change in school inputs has little effect on his decisions.

Figure 5 helps in further illustrating the impact of the school environment on type L and type H. The left graph in each panel plots on the x-axis the cognitive skills that type L would achieve in each school if he successfully graduates, and a on the y-axis a measure of attainment (graduation in panel (a) and high school enrollment in panel (b)). The right graph replicates the same exercise for type H.

While the variation of cognitive cognitive skills across schools is similar for the two types, for type H the probability of graduation is above 94% in all schools and close to 1 in all schools with above average expected cognitive skills; the probability of enrolling in high school conditional on graduation is always above 80% and above 90% in most schools. For type L, attainment exhibit large dispersion across schools: the probability of graduation ranges from 55% to 90% and the conditional probability of enrolling in high school from 30% to 75%. The correlation between skills and attainment is positive but not high: many schools that grant an average level of skills ensure higher attainment than other schools that have a much larger effect on skills. Improving cognitive skills is one of the channel through which schools can increase the attainment of their students, but it is surely not the only one, and perhaps it is not the most relevant.

For type H, the x-axis alone conveys the most salient information to evaluate how attending a given school would impact him. Conversely, for type L, multiple dimensions are necessary to understand how a given school would contribute to his educational outcomes. 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, 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 affect much his prospects. Conversely, the school environment can determine dramatic changes for type L attainment.

Table A-14 in the appendix replicates the analysis discussed in this section with peer variables set at their average value in the sample and only school inputs which change across schools. Results confirm that there would be large differences across schools in the outcomes of type L even if peers were evenly distributed.

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 has also 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 peer quality and school inputs 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.

Panel a) of Table 11 shows result for type L student, using the average school environment among students with low educated parents as in column (1) of Table 9.[53] Retention at time 1 increases cognitive skills at the end of first level by almost 0.2 s.d. Retention at

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

time 2 for students enrolled in second level increases cognitive skills at the end of second level by almost 0.5 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 53%, while for an identical student regularly promoted to next level it is 88%; moreover the *ex-ante* probability of enrolling in high school for the retained student is 14%, while it is about 50% for an identical student who did not experience retention. Retention more than double the odds of dropping out at time $t = 1$. Even if the student stays for one more period, he has 23% probability of dropping out at time $t = 2$. Similarly, a student who is retained in level II has a sizable probability of dropping out, despite very high chances of graduating the following year if he stays. Conditional on graduation, retained students have low probability of enrolling in high school (37% for those who repeated level II, and 27% for those who repeated level I), despite the increase in cognitive skills.

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. This can happen both because retained students like less to be in school or because they have better outside options in the labour market because they are older (or a combination of the two).

Second, as shown in the last lines of Table 11 which compare real and perceived cognitive skills, retained students have lower beliefs than identical students who were promoted, although differences are not large in size. At time $t = 1$ the true level of cognitive skills for type L is -0.46, but the average perceived value for a retained student is lower (-0.52), while it is higher for a promoted student (-0.42). Similarly at time $t = 2$, the true value for a student in second level is -0.46, but on average student who graduate have a perceived value of -0.42 while those who are retained in second period on average believe that it is -0.54. 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.[54]

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

[54]Table A-17 in the appendix replicates 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. Moreover his probability of enrolling in high school after graduation is largely impacted, because it shrinks from 90% to less than 60% if he is retained in first level. It is, however, worth to recall that retention is a concern for a relatively small fraction of type H students (about 3% of them are retained in first level and only 1% in second level).

unobserved. Results in Table A-15 show that removing uncertainty about ability would increase type L graduation probability by about 2 p.p. This is mainly due to an increase in graduation rate (by 3 p.p.) of retained students, who are less likely to dropout both at time $t = 1$ and at $t = 2$. Conversely, overall the rate of enrollment in high school would be almost unchanged: in fact it slightly decreases for students who graduate at $t = 2$ (given that they somewhat overestimate their ability) and it slightly increases for those who graduate at $t = 3$. 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, panel B of Table 11 studies the variation of type L outcomes across schools by retention status. Statistics are computed using the simulation in Table 10. Retention not only decreases graduation rate in each school, but it also amplifies differences across schools: any change in school inputs has a larger effect on type L probability of dropping out when he is retained, because his utility is closer to 0. When retained, type L has a generally low probability of enrolling in high school, however, conditional on graduation, the variance across schools is similar for all retention status.

6.4 Counterfactual school inputs and compulsory education law

I now use the estimated parameters to simulate students' choices and outcomes under counterfactual scenarios. In the first counterfactual, education is mandatory until the end of lower secondary education, rather than being compulsory until a given age. More specifically, students are not allowed to dropout at $t = 1$, and retained students have to stay for three periods, they cannot leave at $t = 2$. Those who fail at $t = 3$ receive a certificate of completion but, as in the baseline scenario, only those who properly graduate can access upper secondary education. Results are shown in column "No Drop." of Table 12.

In the other counterfactuals, selected school inputs are improved up to a given threshold. These exercises can be regarded as simulations of government interventions that target schools lagging behind in one or more dimensions, and that act on school resources and personnel beyond peers, keeping peer quality constant. For instance, improving school inputs 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. Moreover, school inputs 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 about the broader opportunities they would gain if they acquire a high school diploma. Grading and retention policies may become more homogeneous across schools if differences between internal evaluations and region-wide tests are monitored and principals are made accountable. These exercises abstract from the costs that the interventions would entail, but allow to quantify outcomes of interventions that would involve a given number of schools with the goal of improving one or more dimensions.

First, I raise inputs on cognitive skills for schools in the lowest three quartiles: their school input $\mathcal{A}$ is replaced with the value of the highest quartile (column "$\uparrow \mathcal{A}$"). Then, I raise inputs which affect choices and progression in school. In column "$\uparrow \mathcal{M}$", I raise school inputs $\mathcal{M}_M$ and $\mathcal{M}_A$ on choices: lower inputs are replaced with the value at the 75 percentile. School leniency is increased in a similar way in column "$\uparrow \mathcal{I}$"; a larger $\mathcal{I}$ makes retention less likely and has a direct positive effect on choices.[55] Finally, I raise simultaneously all school inputs ($\mathcal{A}$, $\mathcal{M}_M$, $\mathcal{M}_A$, $\mathcal{I}$) in the lowest tertile at the 33 percentile value. While the previous exercises simulate interventions that largely boost a specific dimension for most schools, this last exercise simulate an intervention aimed to improve the most severe deficiencies in all dimensions, raising school inputs at a better – but still below average – value.

Panel a) of Table 12 summarizes the average outcomes in the population. As in previous exercises, I simulate 1000 copies of each individual and, for each of copy, I draw ability and shocks to evaluations, preferences, and retention events, using the estimated distributions; I can then compute outcomes, cognitive skills beliefs, and choices, using the baseline or the counterfactual parameters. Table A-18 in the Appendix shows the average outcomes for the three subgroups of the population defined by parental education. Panel b) of Table 12 focuses on the expected outcomes for type L student. As in previous Subsection 6.2, I simulate the outcomes of 10000 copy of type L student in each school.[56]

Changing the rules for mandatory education would largely increase graduation rate in the population (up to about 94%); in fact most students are able to graduate if they stay in schools until time $t = 3$. The increase is particularly large for type L students, and more in general for students with low educated parents (+17 p.p.), but, as shown in Table A-18, it is also sizeable for students with average parental background (+10 p.p.). Thanks to the large increase in the pool of graduate students, enrollment in high school increases by 4 p.p. overall, while enrollment conditional on graduation slightly decreases. In fact, students at the margin of dropping out are not very likely to pursue further academic education if they eventually graduate. Raising school inputs that directly or indirectly

[55] Recall from Section 5.3.1 that the total school input $\mathcal{T}_M$ (or $\mathcal{T}_A$) can be decomposed as linear combination of $\mathcal{I}$ and $\mathcal{M}_M$ (or $\mathcal{M}_A$).

[56] Results in panel b) of Table 12 are an unweighted average across schools. An alternative approach would be to weight outcomes depending on the probability of type L of being in each school. I replicated the table using as weights the share of students with low parental education enrolled in the school, and results are fully aligned. Graduation rate and enrollment in high school are slightly lower, but comparison of baseline and counterfactuals is extremely similar using weights.

affect educational choices has a smaller impact on graduation rate, which is raised up to 89%, but it is more effective to improve enrollment in high school after graduation (from 77% to up to 82%, with larger growth for students with less educated parents).

The various interventions to raise school inputs at the 75 percentile have similar effects on the average graduation rate and enrollment in further academic education, although they involve different schools. Again, students with low parental background would experience the largest improvements, but changes would be sizeable also for students with average parental background.

The last counterfactual, which improve the worst school inputs in all dimensions, delivers slightly lower but overall similar results, suggesting that targeting schools that lag behind with tailored interventions may also be effective.

By construction, only the interventions that improve $\mathcal{A}$ can raise cognitive skills. Conversely, changing compulsory education law or increasing other school inputs may decrease the average cognitive skills among graduate students, if they avoid the dropout of the worst performers. In fact, a potential issue for interventions aimed at improving other aspects than cognitive skills is that they might prompt some students to acquire further education even if they do not have the necessary competences to succeed. Results in Table 12 suggests that this should not be a major concern. Improving school inputs on motivation (column “”‘$\uparrow \mathcal{M}$”) or school leniency (column “$\uparrow \mathcal{I}$”) only slightly decreases average cognitive skills of students enrolled in high school, by at most 0.06 s.d. The drop is even less for the larger subgroup of graduate students. Changing compulsory education law has a similar effect on high schoolers’ average skills, while the decrease is somewhat larger among graduate students (about 0.12 s.d.). Overall results are a further confirmation that many children who leave education would have pursued further studies if they experienced a different school environment.

7 Conclusions

Suppose that a policy maker wants to identify the “most successful” schools, in order to investigate their methodology, learn their best practices, and apply them in other schools. The school inputs 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 schools 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 further education, 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 performance is not a harmless assumption. At most, it might be a viable simplification when focusing on students with favorable socioeconomic conditions 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.

The methodological approach proposed in this paper allows to disentangle school effects on attainment through cognitive skills from school grading policies, and school contributions to educational choices. Opening the "black box" of school inputs and understanding the mechanisms that lead to the differentiation across schools are relevant extensions that I leave for future research.

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.

Ammermueller, A. and J. Pischke (2009). Peer effects in European primary schools: Evidence from the progress in International Reading Literacy Study. *Journal of Labor Economics 27*(3), 315–348.

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. R. Walters (2016). Stand and deliver: Effects of boston's 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.

Beuermann, D., C. K. Jackson, L. Navarro-Sola, and F. Pardo (2019, October). What is a good school, and can parents tell? evidence on the multidimensionality of school output. Working Paper 25342, National Bureau of Economic Research.

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 (2019a). Grading on a curve: When having good peers is not good. *Economics of Education Review 73*, 101916.

Calsamiglia, C. and A. Loviglio (2019b, Jul). Maturity and school outcomes in an inflexible system: evidence from catalonia. *SERIEs*.

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.

Giustinelli, P. and N. Pavoni (2017). The evolution of awareness and belief ambiguity in the process of high school track choice. *Review of Economic Dynamics 25*, 93 – 120. Special Issue on Human Capital and Inequality.

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 non–test 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 1992–2016: '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.

8 Tables

Table 1: Descriptive statistics by subgroups of the population

	N	%	eval. PS	eval. MS	drop. at 16	graduate	high school	peer quality	neighbor. SES	first choice
ALL	5140	1.000	0.002	0.160	0.086	0.832	0.656	0.000	-0.000	0.925
low parental edu.	1315	0.256	-0.547	-0.383	0.155	0.697	0.430	-0.514	-0.388	0.919
avg parental edu.	2029	0.395	-0.081	0.043	0.097	0.812	0.621	-0.061	-0.046	0.916
high parental edu.	1796	0.349	0.497	0.575	0.022	0.954	0.860	0.445	0.336	0.940
male	2663	0.518	-0.026	0.175	0.096	0.802	0.602	-0.024	-0.001	0.922
female	2477	0.482	0.031	0.144	0.075	0.865	0.713	0.026	0.001	0.929
Spanish	4353	0.847	0.111	0.252	0.064	0.865	0.692	0.112	0.035	0.931
immigrant	787	0.153	-0.606	-0.484	0.207	0.652	0.456	-0.619	-0.196	0.895
regular	3710	0.722	0.277	0.320	0.032	0.968	0.832	0.158	0.084	0.932
retained in primary	417	0.081	-0.858	-0.784	0.230	0.499	0.266	-0.546	-0.166	0.897
retained in grade 1-3	794	0.154	-0.730	-0.493	0.286	0.406	0.140	-0.394	-0.242	0.903
retained in grade 4	219	0.043	-0.380	-0.445	0.000	0.708	0.283	-0.204	-0.236	0.950

Note. The table reports summary statistics for the sample of students used to estimate the structural model described in the paper. It consists of students who enrolled in a public middle school in Barcelona (Spain) in 2009 or in 2010. "eval. PS" are region-wide tests at the end of primary school, "eval. MS" are region-wide tests at the end of middle school (computed on the subsample who reached the last grade), "first choice" takes value 1 if the student is enrolled in the school at the top of her application list.

Table 2: Descriptive statistics by schools

	high p.e.	Spanish	eval. PS	graduate	high school	eval. MS	neigh. SES	first choice
p10	0.04	0.65	-0.73	0.68	0.45	-0.58	-1.09	0.72
p25	0.14	0.77	-0.33	0.73	0.52	-0.23	-0.74	0.90
median	0.31	0.82	-0.04	0.81	0.63	0.09	-0.07	0.95
p75	0.44	0.91	0.20	0.89	0.73	0.36	0.52	0.98
p90	0.56	0.95	0.39	0.94	0.81	0.52	0.79	1.00

Note. This table reports summary statistics for the 47 public middle schools in Barcelona which are used to estimate the structural model discussed in this paper.

Table 3: Variance of unobserved ability

	$\mu_l^2\widehat{\sigma}$	$\text{Var}(C_l)$	%	$\text{Var}(\text{E}_l(h))$	$\text{Var}(\text{E}_l(C_l))$
$l = 0$	0.280 (0.013)	0.572 (0.019)	49.0 (1.5)	0.112 (0.010)	0.406 (0.017)
$l = \text{I}$	0.576 (0.018)	1.202 (0.068)	47.9 (2.4)	0.443 (0.018)	1.069 (0.084)
$l = \text{II}$	0.678 (0.023)	1.164 (0.039)	58.2 (1.6)	0.545 (0.021)	0.962 (0.036)

Panel A: The first column contains the variance of cognitive skills due to unobserved ability (by row: before starting middle school, in level I, and in level II). The second column contains the total variance of cognitive skills, and the third column the share of total variance due to unobserved ability, i.e. $\mu_l^2\widehat{\sigma}/\text{Var}(C_l)$. Similarly, the fourth column contains the variance of beliefs about unobserved ability, and the fifth column the variance of beliefs about cognitive skills. Bootstrap standard errors in parentheses.

Time	Signals received from 0 to t	Posterior Variance
$t = 0$	r_0	0.1685 (0.0040)
$t = 1$	$r_0, g_{\text{I},1}$	0.0650 (0.0040)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{I},2}$	0.0402 (0.0029)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{II},2}$	0.0392 (0.0028)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}$	0.0321 (0.0019)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{I},2}, g_{\text{II},3}$	0.0286 (0.0022)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{I},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0246 (0.0016)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, g_{\text{II},3}$	0.0281 (0.0021)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}, g_{\text{II},3}$	0.0243 (0.0016)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0243 (0.0016)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0213 (0.0013)

Panel B: Posterior variance for each set of signals at a given time. Bootstrap standard errors in parentheses.

Table 4: Estimates of evaluations parameters.

	Skills			γ
	β_0	β_{I}	β_{II}	
Female	0.034 (0.026)	-0.025 (0.126)	0.100 (0.099)	0.396 (0.047)
Immigrant	-0.256 (0.031)	-0.324 (0.183)	-0.296 (0.136)	0.111 (0.064)
Mother education average	0.241 (0.032)	0.345 (0.131)	0.364 (0.133)	-0.098 (0.057)
Mother education high	0.470 (0.033)	0.686 (0.170)	0.661 (0.141)	-0.102 (0.052)
Father education average	0.231 (0.033)	0.213 (0.124)	0.258 (0.118)	-0.047 (0.045)
Father education high	0.368 (0.038)	0.346 (0.117)	0.375 (0.121)	-0.041 (0.060)
Day of birth	0.247 (0.050)	0.144 (0.064)	0.138 (0.056)	-0.010 (0.047)
Retained in primary school	-0.551 (0.052)	-0.850 (0.091)	-0.738 (0.103)	0.520 (0.070)
Neighborhood SES	0.089 (0.018)	0.056 (0.031)	0.081 (0.027)	0.021 (0.024)
School not first choice	-0.004 (0.046)	0.051 (0.075)	0.040 (0.068)	0.019 (0.056)
Repeat level		0.174 (0.281)	0.557 (0.165)	-0.189 (0.258)
Peer quality		0.288 (0.069)	0.158 (0.046)	-0.270 (0.049)
Share female		0.041 (0.024)	0.005 (0.016)	-0.046 (0.020)
School input ($\mathcal{A}$ or $\mathfrak{I}$): p75 - p25		0.367 (0.087)	0.364 (0.091)	0.408 (0.070)
School input: p80 - p20		0.444 (0.099)	0.441 (0.100)	0.530 (0.075)
School input: st. dev.		0.275 (0.058)	0.273 (0.060)	0.298 (0.039)
Female X school input		-0.035 (0.054)	-0.051 (0.039)	-0.002 (0.041)
Immigrant X school input		0.005 (0.084)	-0.002 (0.060)	-0.025 (0.049)
Mother edu. avg. X school input		-0.031 (0.050)	-0.044 (0.053)	-0.030 (0.033)
Mother edu. high X school input		-0.022 (0.068)	-0.002 (0.059)	-0.009 (0.039)
Father edu. avg X school input		0.030 (0.052)	0.002 (0.049)	0.023 (0.034)
Father edu. high X school input		0.045 (0.052)	0.029 (0.055)	0.012 (0.045)
Previous time-varying regressors		-0.014 (0.051)	0.211 (0.317)	
Unobserved ability	1	1.434 (0.040)	1.555 (0.046)	
ρ_{g_τ}		0.217 (0.010)	0.240 (0.012)	
ρ_{r_τ}	0.423 (0.014)		0.280 (0.011)	

*Note.*The estimation includes cohort dummies effects, two dummy variables that take value one if information on mother or father is missing, and a vector of dummies for primary schools attended. For each peer variable a cubic polynomial is used; the table reports the effect on the average student in the sample of having peers at the mean rather than 1 s.d. below the mean. Moreover, school effects are estimated: the table reports the interquantile range, the difference between the 80 and the 20 percentiles, and the standard deviations (computed weighting the school effect by size of the school). The estimation includes interaction between school effects and the dummies for gender, nationality, parental education. The table reports the change for a given characteristics of an increase of 1 standard deviations in the school effect.

Bootstrap standard errors in parentheses.

Table 5: Estimates of retention parameters

	Repeat grade
Belief cognitive skills	-2.104 (0.089)
in II at $t = 2$	-1.106 (0.104)
In II after repeating I	-0.516 (0.200)
Second time in II	-0.349 (0.293)
$\widehat{C}\times$in II at $t = 2$	-0.384 (0.126)
$\widehat{C}\times$in II after repeating I	0.498 (0.194)
$\widehat{C}\times$second time in II	1.162 (0.419)
Female	-0.796 (0.097)
Immigrant	-0.177 (0.139)
Mother education average	0.089 (0.100)
Mother education high	-0.161 (0.138)
Father education average	-0.007 (0.120)
Father education high	-0.262 (0.157)
Day of birth	-0.122 (0.163)
Retained in primary school	-1.198 (0.292)
Neighborhood SES	-0.030 (0.064)
School not first choice	0.182 (0.180)
Peer quality	0.513 (0.124)
Share female	0.032 (0.053)
$\mathcal{I}$	-2.071 (0.183)
School input: p75 - p25	-0.846 (0.142)
School input: p80 - p20	-1.097 (0.141)
School input: st. dev.	0.617 (0.065)

Note. The estimation includes cohort dummies and two dummy variables that 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 at the mean rather than 1 s.d. below the mean.

Bootstrap standard errors in parentheses.

Table 6: Estimates of choices parameters

	Stay in middle school	Enroll in high school
$\widehat{C}$ (belief cognitive skills)	1.310 (0.220)	1.800 (0.122)
$\widehat{C}\times$second time in I	-0.480 (0.217)	
$\widehat{C}\times$in II after repeating I	-0.128 (0.233)	-0.678 (0.204)
$\widehat{C}\times$second time in II	-0.284 (0.392)	-0.555 (0.410)
Female	0.285 (0.096)	0.691 (0.095)
Immigrant	-0.099 (0.109)	0.588 (0.172)
Mother education average	-0.069 (0.108)	-0.023 (0.117)
Mother education high	0.081 (0.166)	-0.119 (0.161)
Father education average	-0.205 (0.111)	0.248 (0.115)
Father education high	-0.372 (0.181)	0.563 (0.169)
Day of birth	-0.331 (0.155)	-0.165 (0.177)
Retained in primary school	0.350 (0.135)	0.987 (0.205)
Neighborhood SES	-0.130 (0.071)	0.102 (0.084)
School not first choice	-0.022 (0.169)	0.063 (0.178)
Second time in I	-1.250 (0.236)	
Second time in II	0.168 (0.479)	-1.730 (0.313)
In II after repeating I	-0.616 (0.415)	-1.773 (0.176)
Peer quality	-0.251 (0.136)	0.137 (0.136)
Share female	-0.127 (0.073)	-0.061 (0.073)
School input ($\mathcal{T}_M$ or $\mathcal{T}_A$): p75 - p25	0.684 (0.140)	0.904 (0.149)
School input: p80 - p20	0.827 (0.149)	0.996 (0.157)
School input: st. dev.	0.495 (0.059)	0.610 (0.077)
$\mathcal{I}$	1.137 (0.177)	1.009 (0.254)
School input $\mathcal{M}_M$ or $\mathcal{M}_A$: p75 - p25	0.383 (0.119)	0.717 (0.140)
School input $\mathcal{M}_M$ or $\mathcal{M}_A$: p80 - p20	0.604 (0.137)	0.933 (0.185)
School input $\mathcal{M}_M$ or $\mathcal{M}_A$: st. dev.	0.357 (0.054)	0.531 (0.073)

Note. The estimation includes cohort dummies and two dummy variables that 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 at the mean rather than 1 s.d. below the mean. School effects are estimated: the table reports the interquantile range, the difference between the 80 and the 20 percentiles, and the standard deviation (computed weighting the school effect by size of the school) of the estimated school effects.

Bootstrap standard errors in parentheses.

Table 7: Fit of the model

	Stay at $t = 1$		Graduate		High school		Retained in I		Retained in II	
	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
ALL	91.44	91.97	83.23	84.21	65.56	65.43	23.56	23.50	4.26	5.84
male	90.42	90.44	80.17	80.76	60.23	59.61	26.96	27.14	4.84	6.62
female	92.53	93.61	86.52	87.91	71.30	71.70	19.90	19.58	3.63	4.99
Spanish	93.64	93.76	86.49	87.27	69.17	68.78	18.91	19.13	4.11	5.64
immigrant	79.29	82.07	65.18	67.28	45.62	46.95	49.30	47.65	5.08	6.92
low parental edu.	84.49	85.04	69.73	70.04	43.04	42.74	43.19	42.32	6.62	8.59
avg parental edu.	90.29	91.41	81.17	83.13	62.10	62.48	25.68	25.60	4.88	6.65
high parental edu.	97.83	97.67	95.43	95.80	85.97	85.39	6.79	7.34	1.84	2.91
below median peers	86.64	88.51	75.88	77.77	53.40	54.36	33.24	32.21	5.07	6.78
above median peers	94.89	94.46	88.50	88.83	74.30	73.38	16.61	17.24	3.68	5.16

Note. Data frequencies are computed from the sample of students used in the estimation. The model frequencies are constructed using 1000 simulations of the structural model for each individual included in the estimation.

Table 8: Average school input by parental education

	$\mathcal{A}$	$\mathcal{I}$	$\mathcal{T}_M$	$\mathcal{T}_A$	$\mathcal{M}_M$	$\mathcal{M}_A$
low parental edu.	0.019	0.014	0.005	0.251	-0.007	0.280
high parental edu.	-0.032	0.021	-0.035	-0.188	-0.069	-0.228
p-value difference	0.15	0.847	0.273	1.3e-33	0.0981	2.26e-45

Note. The first line of the table displays average school inputs for students with low educated parents. The second line shows average school inputs for students with highly educated parents. The third line contains the p-value of a t-test for the difference of the means. The estimated school inputs are standardized to have mean 0 and standard deviation 1 in the sample.

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)
graduate	0.922	0.774	0.757	0.805	0.788	0.990	0.993	0.985	0.988
enrol in high school	0.705	0.396	0.338	0.490	0.433	0.928	0.944	0.891	0.915
$C_{\mathrm{I},1}$	0.174	-0.456	-0.470	-0.198	-0.213	0.871	0.886	0.613	0.628
repeat level I	0.145	0.306	0.311	0.289	0.293	0.032	0.031	0.035	0.034
dropout at $t = 1$	0.042	0.112	0.125	0.083	0.095	0.007	0.005	0.012	0.009
drop at $t = 2$	0.121	0.205	0.218	0.171	0.186	0.044	0.033	0.061	0.050
graduate|stay	0.984	0.949	0.948	0.940	0.939	0.999	0.999	0.999	0.999
enrol in h.s.|grad.	0.765	0.512	0.446	0.609	0.549	0.937	0.951	0.905	0.926
C_0	0.069	-0.336	-0.336	-0.336	-0.336	0.502	0.502	0.502	0.502
C_{II}|grad.	0.170	-0.398	-0.412	-0.193	-0.207	0.848	0.863	0.642	0.658
r_{II}	0.110	-0.350	-0.361	-0.184	-0.195	0.659	0.672	0.493	0.505
g_{II}	-0.253	-0.569	-0.581	-0.634	-0.645	0.282	0.295	0.346	0.360

Note. Frequencies are constructed using 10000 simulations of the structural model for each student type. Type L has parents with basic education; type H has parents with tertiary education. The first column of the table reports outcomes for a student with average parental background. The fictitious students created in the simulation are male, Spanish, born on July 1, they began middle school at 12 years old, they live in a neighborhood with average quality, 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 similar 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		
	p80-p20	p75-p25	s.d.	p80-p20	p75-p25	s.d.
graduate	0.152	0.123	0.084	0.010	0.007	0.008
enrol in high school	0.171	0.116	0.098	0.055	0.041	0.036
$C_{\mathrm{I},1}$	0.596	0.485	0.402	0.635	0.492	0.418
repeat level I	0.122	0.104	0.078	0.020	0.016	0.015
dropout at $t = 1$	0.075	0.068	0.047	0.007	0.006	0.006
drop at $t = 2$	0.092	0.077	0.053	0.026	0.022	0.016
graduate|stay	0.060	0.049	0.033	0.001	0.001	0.001
enrol in h.s.|grad.	0.166	0.123	0.100	0.048	0.035	0.032
C_{II}|grad.	0.500	0.430	0.358	0.535	0.447	0.383

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.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 peer characteristics in each school are used.

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

(a) Average school environment in the sample

	not retained	retained in I	retained in II
dropout at t=1	0.0843	0.1749	-
graduate (Pr at t=1)	0.8848	0.5247	-
enrol in high school (Pr at t=1)	0.5083	0.1430	-
graduate at t=2|stay at t=1	0.8440	-	-
enrol in hs at t=2|grad. at t=2	0.6039	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.2347	0.1279
graduate at t=3|stay at t=2	-	0.8309	0.8981
enrol in hs at t=3|grad. at t=3	-	0.2726	0.3715
true $C_{\mathrm{I},1}$	-0.4561	-0.4561	-
perceived $\widehat{C}_{\mathrm{I},1}$	-0.4248	-0.5252	-
true $C_{\mathrm{I},2}$	-	-0.2817	-
perceived $\widehat{C}_{\mathrm{I},2}$	-	-0.3041	-
true $C_{\mathrm{II},2}$	-0.4617	-	-0.4617
perceived $\widehat{C}_{\mathrm{II},2}$	-0.4157	-	-0.5426
true C_{II}|graduate	-0.4617	-0.4250	0.0957
perceived $\widehat{C}_{\mathrm{II}}$|graduate	-0.4157	-0.4126	0.0528

Note. Statistics are computed using the simulation in column (1) of Table 9, for Type L (low educated parents). Peers characteristics and school inputs take the average values among students with low parental background.

(b) Variation across schools

	Graduate		High school		High school | grad.	
	p75-p25	s.d	p75-p25	s.d	p75-p25	s.d
promoted at t=1	0.065	0.042	0.126	0.100	0.123	0.100
retained at t=1	0.157	0.097	0.063	0.058	0.113	0.092
graduate at t=2	-	-	-	-	0.122	0.099
retained at t=2	0.107	0.070	0.106	0.090	0.125	0.103

Note. Statistics are computed using the simulation in Table 10, for Type L (low parental education). Column display the interquantile range and the standard deviation across schools for graduation rate, enrollment in high school, and enrollment in high school conditional on graduation.

Table 12: Counterfactual outcomes

(a) Average outcomes in the population

	Baseline	Counterfactuals				
		No drop.	$\uparrow \mathcal{A}$	$\uparrow \mathcal{M}$	$\uparrow \mathcal{I}$	$\uparrow$ bottom
graduate	0.842	0.938	0.875	0.867	0.890	0.870
enrol in high school	0.654	0.691	0.707	0.710	0.712	0.691
enrol in h.s.|grad.	0.777	0.737	0.808	0.819	0.800	0.795
$C_{\mathrm{II}}\|$ grad.	0.254	0.129	0.394	0.224	0.197	0.262
$C_{\mathrm{II}}\|$ h.s.	0.464	0.402	0.584	0.408	0.400	0.463

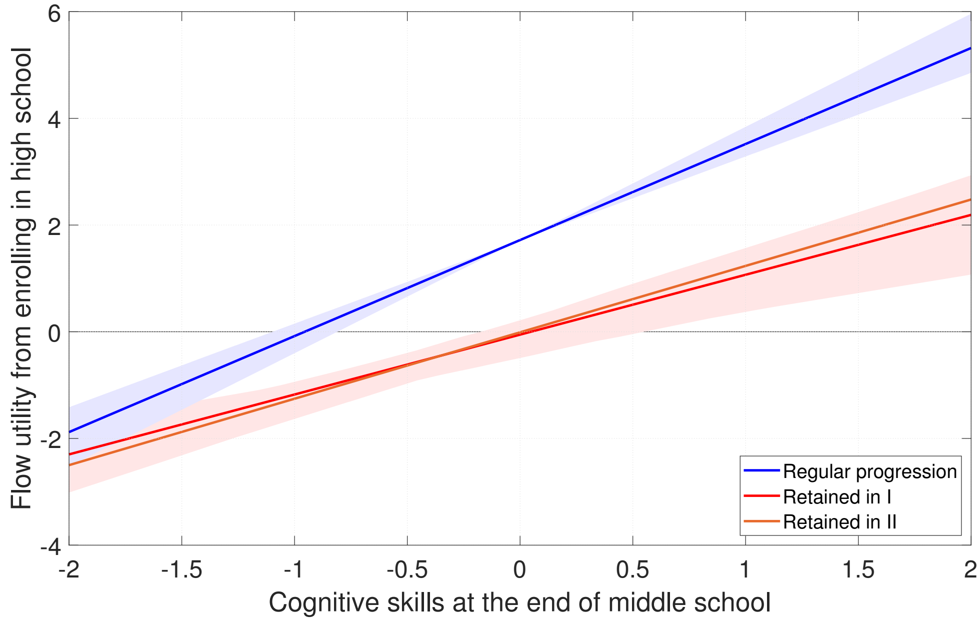
(b) Average outcomes for type L

	Baseline	Counterfactuals				
		No drop.	$\uparrow \mathcal{A}$	$\uparrow \mathcal{M}$	$\uparrow \mathcal{I}$	$\uparrow$ bottom
graduate	0.770	0.922	0.841	0.817	0.857	0.825
enrol in high school	0.402	0.457	0.511	0.493	0.507	0.470
enrol in h.s.|grad.	0.519	0.495	0.601	0.598	0.586	0.565
$C_{\mathrm{II}}\|$ grad.	-0.369	-0.363	-0.181	-0.367	-0.383	-0.327
$C_{\mathrm{II}}\|$ h.s.	-0.387	-0.380	-0.195	-0.381	-0.396	-0.342

Note. Average outcomes in the column "Baseline" are computed using the estimated parameters of the model. The other columns contain results of counterfactual simulations. In "No drop.", students are not allowed to dropout. In column "$\uparrow \mathcal{A}$", school inputs on cognitive skills are replaced with the value at the 75 percentile if they are lower. Similarly for school inputs $\mathcal{M}_M$ and $\mathcal{M}_A$ on choices, and school leniency in $\mathcal{I}$ in the following columns "$\uparrow \mathcal{M}$" and "$\uparrow \mathcal{I}$" respectively. In column "$\uparrow$ bottom", all school inputs in the bottom tertile are replaced with the value at the 33 percentile. In panel a) values are the average of 1000 simulations of the structural model for each individual in the sample. In panel b) values are the average of 10000 simulations for type L student in each school in the sample. The average peer variables in each school are used.

9 Figures

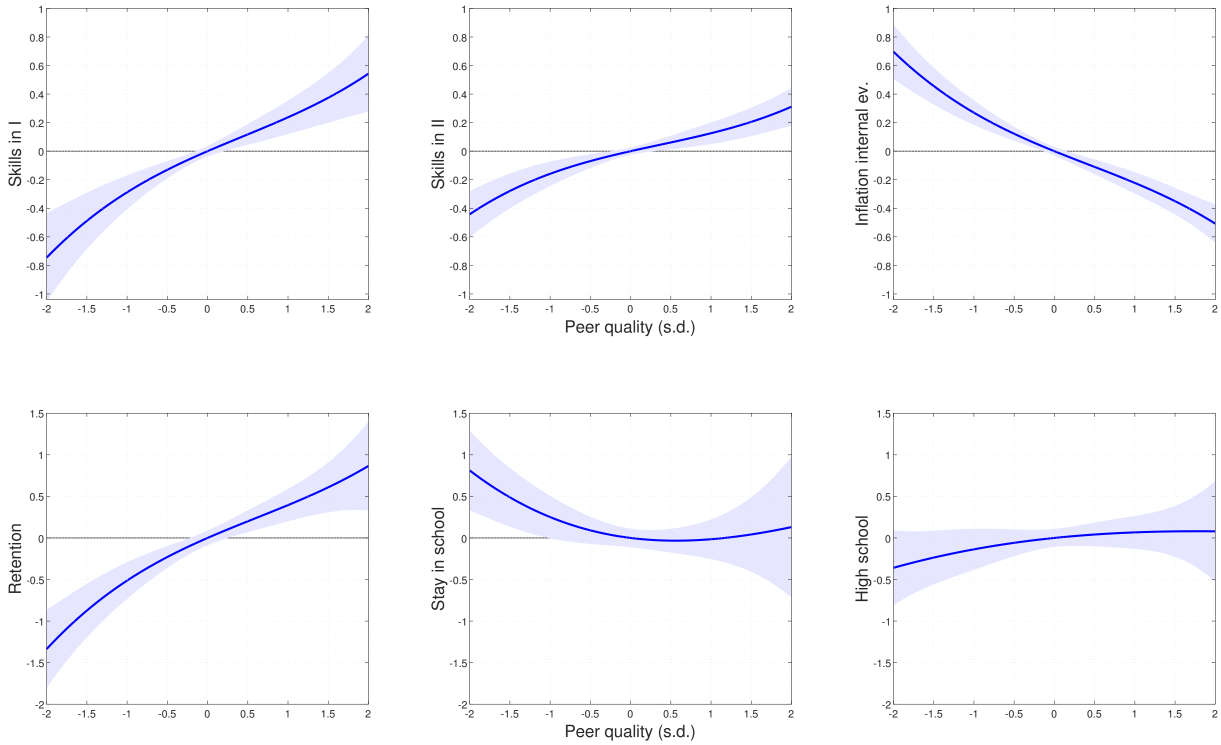
Figure 1: Effect of cognitive skills on high school enrollment by retention status



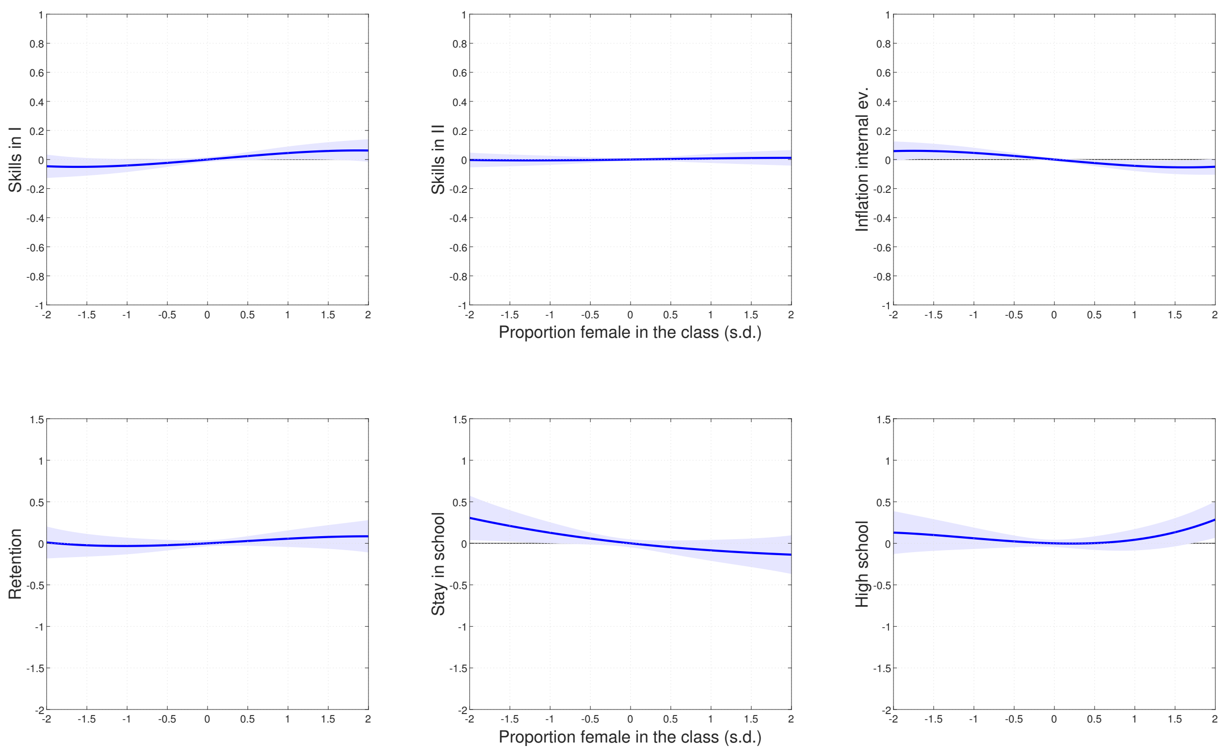
Note. The figure plots the effect of beliefs about cognitive skills on the utility from the choice of enrolling in high school for three types of students: those who completed middle school regularly at time $t = 2$ (blue line), those who graduated at time $t = 3$ because they were retained at time$t = 1$ and repeat level I (red line), those who graduated at time $t = 3$ because they were retained at time $t = 2$ and repeat level II (orange line). Shaded areas are 95% confidence intervals. The confidence interval for "Retained in II" is not plotted for clarity, it largely overlaps with the one for "Retained in I".

Figure 2: Effects of peer variables on outcomes

(a) Peer quality

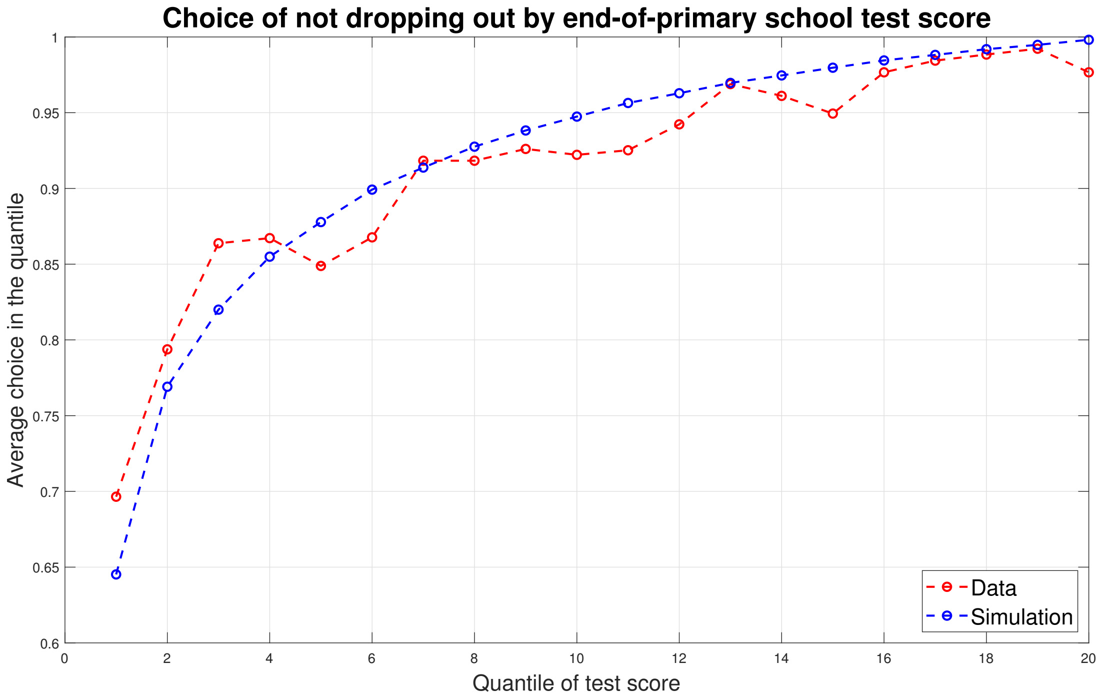


(b) Share of female



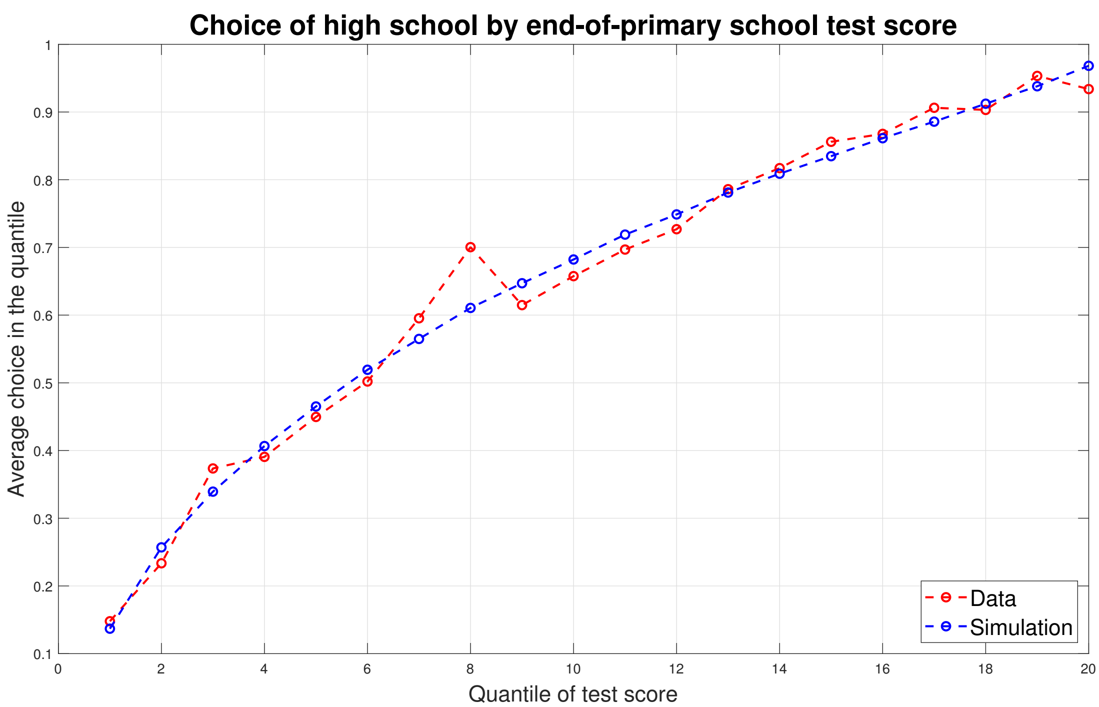
Note. The figures plot the effects of peer quality (panel (a)) and share of female (panel (b)). Shaded areas are 95% confidence intervals.

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.

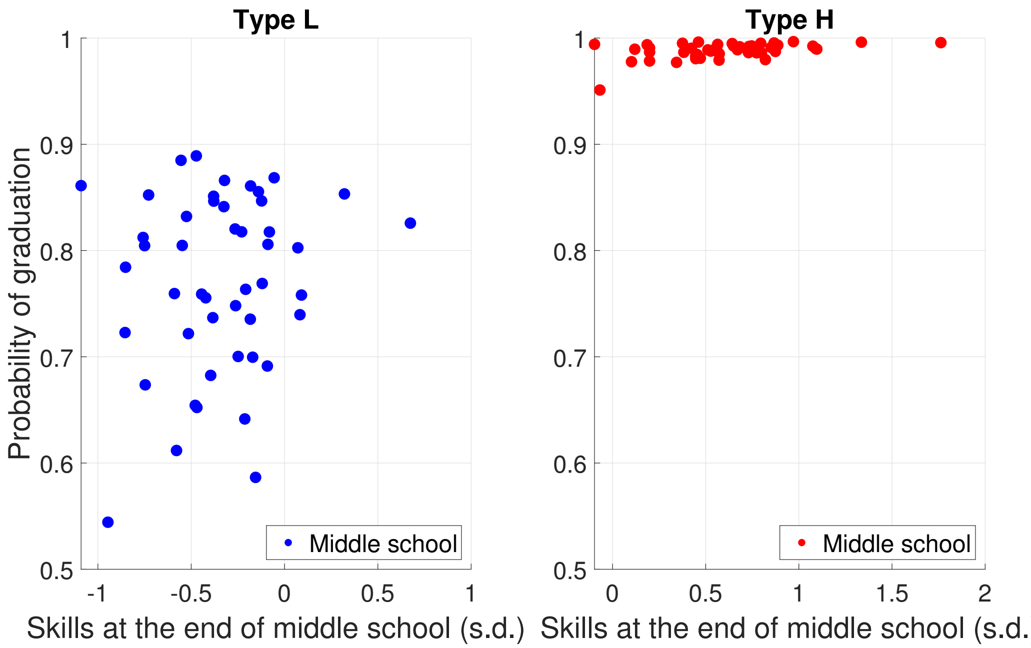
Figure 4: Fit of the model (ii)



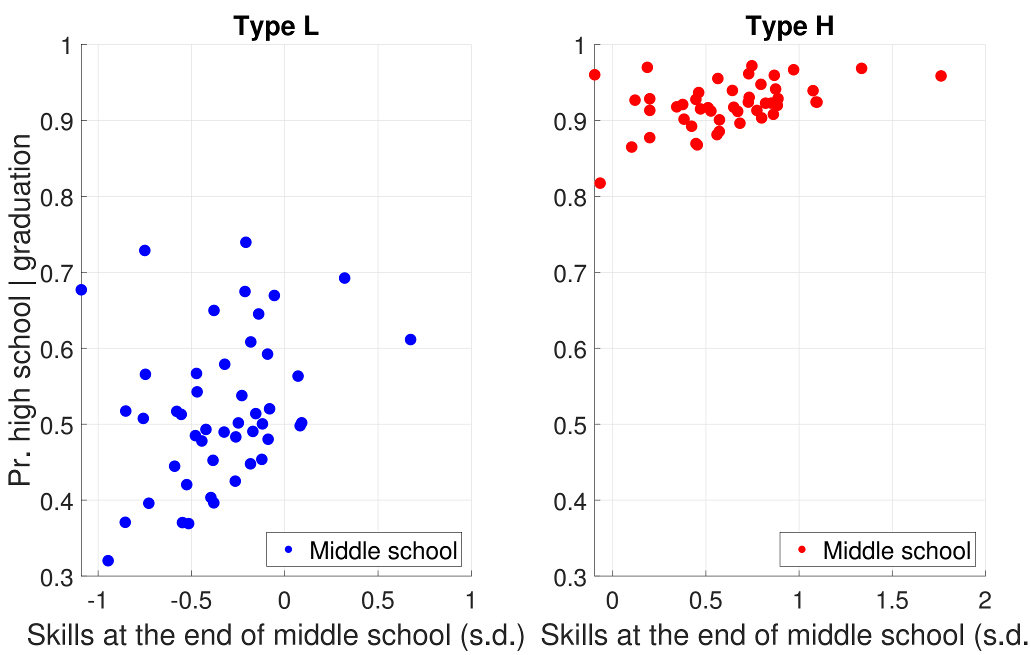
Note. The figure plots the share of students who enroll in high school by quantile of their test score at the end of primary school. Sample average from the real data are in red, while results of the simulation performed using the estimated parameters of the model are in blue.

Figure 5: Expected outcomes by school

(a) Graduation



(b) Enrollment in high school



Note. Each dot in the figures plots predicted educational outcomes for a student with low educated parents (blue dots, left) and for a student with highly educated parents (red dots, right) in a given school. The x-axis plots the cognitive skills that he would acquire if he graduates, the y-axis plots the probability of graduation (panel (a)) or the probability of enrolling in high school, conditional on graduation (panel (b)). Expected outcomes at the school level are computed using data of the simulation described in Section 6.2 and Table 10.

Appendixes for online publication

A Additional tables

Table A-13: Fit of the model (bis)

	graduate at $t = 2$		graduate at $t = 3$		enroll in hs	
	Data	Model	Data	Model	Data	Model
ALL	94.25	92.05	83.43	85.27	78.78	77.71
male	93.17	90.46	80.30	82.67	75.13	73.81
female	95.32	93.57	87.57	88.80	82.41	81.55
Spanish	94.83	92.77	83.07	85.37	79.97	78.81
immigrant	88.54	85.58	84.30	84.97	69.98	69.78
low parental edu.	87.66	83.82	84.70	82.41	61.72	61.02
avg parental edu.	93.15	90.62	82.42	85.96	76.50	75.16
high parental edu.	98.01	96.81	82.69	90.98	90.08	89.13
below median peers	91.97	89.36	84.51	85.22	70.37	69.90
above median peers	95.52	93.58	82.11	85.33	83.95	82.61

Note. Data frequencies are computed from the sample of students used in the estimation. The model frequencies are constructed using 1000 simulations of the structural model for each individual included in the estimation. Statistics are conditional on reaching the relevant level, for instance graduation at $t = 2$ is conditional on being promoted at $t = 1$ and choosing to stay in education.

Table A-14: 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
graduate	0.768	0.145	0.126	0.084	0.987	0.011	0.007	0.008
enrol in high school	0.436	0.174	0.126	0.126	0.920	0.047	0.042	0.036
$C_{\mathrm{I},1}$	-0.315	0.455	0.375	0.290	0.770	0.493	0.406	0.313
repeat level I	0.308	0.120	0.102	0.077	0.035	0.021	0.016	0.015
dropout at $t = 1$	0.111	0.071	0.054	0.047	0.009	0.007	0.005	0.006
drop at $t = 2$	0.195	0.089	0.073	0.053	0.064	0.036	0.032	0.023
graduate|stay	0.937	0.053	0.041	0.034	0.998	0.002	0.001	0.001
enrol in h.s.|grad.	0.563	0.191	0.152	0.125	0.932	0.042	0.037	0.032
C_{II}|grad.	-0.271	0.438	0.355	0.278	0.768	0.502	0.405	0.314

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 basic education; type H has parents with tertiary education. 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 neighborhood quality, the average cohort effect and the average primary school effect in the sample. Peer variables are set at the average in the sample.

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	0.0835	0.1507	-
graduate (Pr at t=1)	0.8906	0.5528	-
enrol in high school (Pr at t=1)	0.5005	0.1476	-
graduate at t=2|stay at t=1	0.8567	-	-
enrol in hs at t=2|grad. at t=2	0.5862	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.2223	0.1130
graduate at t=3|stay at t=2	-	0.8370	0.9047
enrol in hs at t=3|grad. at t=3	-	0.2670	0.3815
true $C_{\mathrm{I},1}$	-0.4561	-0.4561	-
true $C_{\mathrm{I},2}$	-	-0.2817	-
true $C_{\mathrm{II},2}$	-0.4617	-	-0.4617
true C_{II}|graduate	-0.4617	-0.4250	0.0957

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	0.121	0.284	0.290	0.266	0.271	0.025	0.024	0.027	0.026
dropout at t=1	0.031	0.103	0.118	0.072	0.084	0.004	0.003	0.008	0.006
drop at t=2|not grad.	0.103	0.192	0.207	0.158	0.172	0.040	0.033	0.054	0.045
graduate	0.941	0.795	0.774	0.829	0.812	0.994	0.995	0.990	0.992
enrol in high school	0.733	0.400	0.338	0.504	0.441	0.938	0.952	0.903	0.926
enrol in hs|graduation	0.779	0.504	0.437	0.608	0.543	0.943	0.956	0.913	0.933

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	0.0060	0.0277	-
graduate (Pr at t=1)	0.9934	0.8986	-
enrol in high school (Pr at t=1)	0.9393	0.5893	-
graduate at t=2|stay at t=1	0.9900	-	-
enrol in hs at t=2|grad. at t=2	0.9470	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.0490	0.0260
graduate at t=3|stay at t=2	-	0.9717	0.9701
enrol in hs at t=3|grad. at t=3	-	0.6558	0.7822
true $C_{\mathrm{I},1}$	0.8707	0.8707	-
perceived $\widehat{C}_{\mathrm{I},1}$	0.8747	0.7669	-
true $C_{\mathrm{I},2}$	-	1.0451	-
perceived $\widehat{C}_{\mathrm{I},2}$	-	0.9913	-
true $C_{\mathrm{II},2}$	0.8414	-	0.8414
perceived $\widehat{C}_{\mathrm{II},2}$	0.8456	-	0.7014
true $C_{\mathrm{II}}\|$graduate	0.8414	0.8782	1.3988
perceived $\widehat{C}_{\mathrm{II}}\|$graduate	0.8456	0.8495	1.3076

Table A-18: Counterfactual outcomes by parental education

	Baseline	Counterfactuals				
		No drop.	$\uparrow \mathcal{A}$	$\uparrow \mathcal{M}$	$\uparrow \mathcal{I}$	$\uparrow$ bottom
Low parental education						
graduate	0.702	0.876	0.757	0.744	0.784	0.751
enrol in high school	0.424	0.480	0.494	0.487	0.503	0.473
enrol in h.s.|grad.	0.604	0.548	0.653	0.655	0.642	0.630
$C_{\text{II}}\|$ grad.	-0.378	-0.546	-0.250	-0.416	-0.438	-0.369
$C_{\text{II}}\|$ h.s.	-0.116	-0.213	-0.013	-0.170	-0.180	-0.112
Average parental education						
graduate	0.831	0.936	0.868	0.859	0.887	0.862
enrol in high school	0.625	0.669	0.683	0.689	0.693	0.668
enrol in h.s.|grad.	0.752	0.715	0.786	0.802	0.782	0.775
$C_{\text{II}}\|$ grad.	0.097	-0.009	0.227	0.073	0.047	0.110
$C_{\text{II}}\|$ h.s.	0.271	0.213	0.384	0.223	0.215	0.276
High parental education						
graduate	0.957	0.988	0.971	0.969	0.972	0.967
enrol in high school	0.853	0.870	0.888	0.895	0.883	0.875
enrol in h.s.|grad.	0.891	0.880	0.914	0.924	0.908	0.905
$C_{\text{II}}\|$ grad.	0.738	0.703	0.923	0.726	0.719	0.765
$C_{\text{II}}\|$ h.s.	0.831	0.811	0.998	0.795	0.804	0.849

Note. This table reports the outcomes of Table 12 by subgroups of parental education. "Low parental education" means that parents have at most lower secondary education. "High parental education" means that at least one parents has a college degree and the other has at least upper secondary education (or the second parent is missing). The remaining observations belong to the subgroup "Average parental education".

B Data Appendix

B.1 Sample Selection

As explained in Section 2, the sample consists of students who enroll for the first time in lower secondary education in September 2009 or September 2010 in a public middle schools in Barcelona, who do not have special education needs (98.5%), and for whom I could retrieve the results in a standardized test at the end of primary education (77%). This subsection details further data cleaning.

I dropped 13 students who were younger than 12 years old or older than 13 when they enroll.[57] I dropped 89 students who appear in the enrollment data only once at 12/13 years old, assuming that they moved in another region, and 45 students for whom I could not retrieve internal evaluations in first period.

Less than 10% of the students change school during lower secondary education. About half of them stay in the public system, while the other switch to a non-public school. I observe enrollment data for the latter, but I don't have information on performance, therefore I drop them from the analysis. Conversely students who change school within the public system are included in the final sample. They are assigned the middle school that they attend for the first year, and the peers that they would have if they staid in the same class.[58]

Finally, I restrict the analysis to the schools with enough students in each cohort, and in which those students are a large share of the children enrolled in first grade. More specifically, I focus on schools with at least 18 students in the sample in each cohort (dropping 5 schools) and with more than 40% of pupils in first grade who belong to the sample (dropping 1 school). Thus, I exploit 47 out of 53 public middle schools that appear in the enrollment data in the time period under analysis.[59]

B.2 Variables

Individual characteristics. As described in Data Section 2.

Neighborhood Socioeconomic Status. The enrollment data contain the postal code of the student's address. I use the postal code zone as proxy for the neighborhood. About 96% of the students in the final sample used in this paper live in Barcelona, which

[57]As explained in Section 2.1, usually students are 12 years old when they enroll (or turn 12 before the end of the year). They are 13 years old if they repeat a year during primary education.

[58]This approach introduces some measurement error, however results are very similar if those observations are omitted from the sample.

[59]The chosen thresholds exclude about 3% of the students in the selected sample. In a previous version of this paper, I imposed more restrictive thresholds and exploited data from only 44 schools. Overall, results are very similar.

is divided in 42 postal code areas. The remaining 4% of students come from nearby municipalities. I create an index based on the average gross income at the postal code level and on the share of highly educated mothers and fathers. More specifically, I use data from the Spanish Tax Agency, which contain information at the postal code level for Barcelona and other large municipalities (Badalona, Hospitalet de Llobregat, Sabadell) and at the town level for the other few small municipalities that appear in the sample. Moreover, in the absence of an external source with detailed information on the education of the resident population, I use the enrollment data in Catalonia (merged with parental education) to compute the shares of college educated mothers and fathers among the parents of children aged 6 to 15 years old who live in the postcode zone in the years under analysis. While those share are only a proxy for the shares of college graduates in the overall population, they refer to the group of adults that may be more relevant for the teens enrolled in schools. I perform a principal component analysis and use as index the first component, which explain 92% of the variance and load similarly on the three factors.[60] The index is normalized to have mean 0 and standard deviation 1 in the sample.

Peer characteristics. Peer variables are defined as the average value among all the peers in the class, including students who do not belong to the sample used for the analysis (for instance because they belong to an older cohort and were retained). While I observe peers at $t = 1$ for everyone, I need to define counterfactual peers at time $t = 2$ and $t = 3$ for those who dropout, and peers at $t = 3$ for those who graduate on time. I use the average values among peers who stay and/or repeat the grade.[61] In the very few cases in which there are no peers in the relevant situation, for instance no one in the class fails and repeat the year, I use peers at the school level.

I use two peer variables in the main analysis: the share of females and an index of "peer quality". The latter combines information on the share of college graduate among mothers and fathers, the share of immigrant students, the average test score at the end of primary education, and the share of students who took the test.[62] I perform pca and use the first component, which explains 73% of the variance. The loading factors are very similar in magnitude ranging from 0.42 to 0.47, the share of immigrant peers has a a

[60] I also replicated the analysis introducing the three variables separately. Results are very similar and coefficients of other regressors are virtually unchanged. Therefore I prefer the more parsimonious specification with one index.

[61] In most cases, students stay in the same class throughout middle schools, but sometimes classes are shuffled, or retained students are assigned to another class, e.g. if there are more free seats there. Thus, such formulation for the counterfactual peers is a weighted average of the groups of peers that children would have had in the counterfactual scenario.

[62] The share of students for whom I could not retrieve the test is also a proxy for recent immigrants who may have limited knowledge of the local languages. In fact they might have been exempted from the exam or they may have completed primary education in another country.

negative signs while the other are positive.[63] Finally, the two variables are standardized to have mean 0 and standard deviation 1 in the sample.

I replicated the analysis using all the peer regressors separately. Overall results are fully aligned but some parameters are not precisely estimated. Therefore I prefer the more parsimonious specification.

C 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:

$$L(r_{i0}; \zeta, \sigma | h) = \frac{1}{\sqrt{2\pi \rho_{r_0}}} \exp\left(- \frac{(r_{i,0} - h - z'_{i,0}\beta_0)^2}{2\rho_{r_0}} \right) \tag{A-3}$$

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

[63] The correlation between "peer quality" and share of female is virtually 0. When I include the share of female in the pca and extract two (rotated) components, the second component is almost identical to the share of female.

(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.[64]

D Students' allocation to classes

While students allocation to schools is centrally regulated (as discussed in Section 2.1), there aren't strict rules about students' allocation to classes within a given school. Anecdotal evidence suggests that in basic education students are typically assigned to classes randomly, in some cases conditionally on their gender. However, it is not possible to rule out that some principals may follow a different approach. Sorting of students across classes might challenge the identification; in particular, the effect of peer quality on performance or choices might reflect unobserved characteristics of the student rather than a proper peer effect.

Following Ammermueller and Pischke (2009), I perform a series of Pearson χ^2 tests to test if students characteristics and the class the students is assigned to are statistically independent. Table A-19 presents the p-values for the main individual characteristics used in the model: gender, immigrant status, mother education and father education. Tests are performed both by cohort and pooling all the observations together. For those

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

variables the null hypothesis of independence is never rejected. On the other hand, the null hypothesis is rejected for the last variable of the table: a dummy for students with above average primary school test score.[65] More specifically, the null hypothesis is rejected in 4 schools for the cohort 2009 and in 6 schools for the cohort 2010. In only 1 school it is rejected for both cohorts. Overall, results do not allow us to fully rule out that in some cases students are assigned to classes taking in account their past performances, although it does not seem that schools systematically sort students.

I estimate the model restricting the sample to the 36 schools for which the null hypothesis of independence is never rejected. The estimated peer effects on choices and retention are close to the baseline specifications. The estimated peer effect on cognitive skills is similar at the tails of the distribution but flatter around the average. This suggests that, if anything, the main specification might slightly overestimate the positive effect on cognitive skills of an improvement of peer quality when the baseline value is close to the average. The estimated school inputs and the coefficients of individual characteristics are quite close to the baseline ones. Moreover, I replicate the simulations discussed in Section 6. Results are qualitatively and quantitatively very similar.

Table A-19: Tests for independence of peer variables and class assignment

	female	immigrant	mother edu.	father edu.	test score
Cohort 2009	0.856	0.546	0.358	0.235	0.000
Cohort 2010	0.991	0.137	0.559	0.355	0.022
All	0.924	0.340	0.458	0.295	0.011

Note. The table reports p-values for Pearson χ^2 tests of independence between the student characteristics and classroom assignment within each school using the individual-level data.

[65] To account for the continuous nature of the variable, I also performed Kruskall-Wallis tests in each school using the primary school test score. Results at the school level are extremely similar, but the aggregation in one statistic is less straightforward than with the Pearson χ^2.