ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

School Quality Beyond Test Scores: the Role of Schools in Shaping Educational Outcomes

Annalisa Loviglio

Quaderni - Working Paper DSE N°1184

School Quality Beyond Test Scores: the Role of Schools in Shaping Educational Outcomes

Annalisa Loviglio*

March, 2023

Abstract

I study how schools impact student performance and educational attainment throughout secondary education, and show that school quality cannot be easily captured by any type of rankings because students with differing characteristics and abilities benefit from different school inputs. To do so, I estimate a dynamic structural model of cognitive skills accumulation and schooling decision using rich administrative data from middle schools in Barcelona. I then simulate the outcomes that each student would have achieved in every school in the sample. Notably, the school environment has a crucial impact on the educational attainment of students from less advantaged family background and low-ability students who are at greater risk of leaving school. Moreover, the schools that would yield the highest final test scores for these students – provided they do not drop out – are not the ones that would maximize their likelihood of graduating and enrolling in further education. The results suggest that evaluating and comparing schools using only standardized assessments is insufficient for serving the needs of disadvantaged students, who require schools that enhance educational attainment rather than just test scores.

KEYWORDS: School Value-Added, Educational Choices, Educational Attainment, Retention
JEL CODES: I20, J24, C35

*Department of Economics, University of Bologna, Italy. Correspondence: annalisa.loviglio@unibo.it. Previously circulated as: "Schools and Their Multiple Ways to Impact Students: A Structural Model of Skill Accumulation and Educational Choices" (first draft November 2018). I am grateful to Caterina Calsamiglia and Joan Llull for their guidance and continuous support during my doctoral studies. I would also like to thank Manuel Arellano, Peter Arcidiacono, Chiara Binelli, Massimilano Bratti, Ana Rute Cardoso, Olivier De Groote, Davide Dragone, Anna Houstecka, Arnaud Maurel, Michele Pellizzari, 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.

Non-technical summary

Providing inclusive and high quality education that improves outcomes for all students, particularly the most disadvantaged, is a primary goal for policy makers all around the world. Therefore, many countries have adopted school accountability measures for the purpose of monitoring school quality, deciding on corrective actions, and in some cases assigning funding. A prime example is the "No Child Left Behind" program in the US. In practice, school quality is often measured using nationwide tests, under the assumption that a school's ability to improve students' test scores is a sufficient measure of the school ability to improve their overall educational outcomes. In this paper, I investigate the perils of this simplification and show that school quality cannot be easily captured by any type of rankings because students with differing characteristics and abilities benefit from different school inputs.

I examine how the school in which a student is enrolled affects their performance, graduation probability, and probability of acquiring further academic education. To do so, I develop a model of cognitive skills accumulation and schooling decisions throughout lower secondary education of students enrolled in heterogeneous schools. In each period, students acquire new skills and decide whether to continue their education or leave school. Skill development and therefore performance depends on the student innate ability, family background and school environment. The decision to drop out depends on the student's level of cognitive skills, but also on non-cognitive factors, such as family and school inputs. Furthermore, before the choice to drop out, a student can be required to repeat a year in order to stay in school, referred to as "retention". Retention may raise skills in the following period, but it also increases the time required to graduate and can therefore alter a student's preferences. Schools differ in their grading and retention policies, thus this is an additional channel through which they can affect students' attainment.

I estimate the model using administrative data on public middle schools in Barcelona (Spain), a setting in which 17% of students do not complete basic education and 35% do not enroll in high school. As in many other countries, Spanish students are legally required to stay in school until they turn 16. Normally, this corresponds to their last year of lower secondary education, but given that retention is common (25% of students in the sample are retained at least once), many of them turn 16 well before their potential graduation date. After graduation, they choose whether to enroll in high school for two more years.

After estimating the parameters of the model, I simulate the outcomes that each student would have achieved in every school in the sample and compute school rankings at the individual level. Specifically, for each student I rank schools based on the student's expected performance at the end of middle school, probability of graduation and probability of enrolling in high school.

On the one hand, results show that both family and school inputs have a large effects on a student's cognitive skills development and, therefore, on their test scores. Furthermore, simulated school rankings by skill level are both very similar across students and similar to the school ranking obtained from the end-of-the-year observed test scores. In other words, schools affect similarly the performance of students with differing parental background, and a simple ranking based on average test scores is a good indicator of how the school contributes to the cognitive skills development of any student, provided that they do not drop out before completion.

On the other hand, schools have a sizable direct effect on educational choices, and therefore on attainment, beyond their indirect effect by way of cognitive skills. Individual school rankings by attainment (i.e. graduation and enrollment in further education) vary across students, are less correlated with the aggregate measures, and are generally different from the individual rankings based on skills. This implies that the commonly used school ranking based on test scores is not informative about a student's educational attainment prospects. Furthermore, other simple statistics at the school level (such as the graduation rate) do not contribute much, because they do not accurately reflect the outcomes that students would have achieved if enrolled there, especially if their individual characteristics differ from the average of students in the school.

Differences in attainment across schools are sizable for the large majority of students with low parental background. For instance, the graduation probability of a representative student with average ability and low parental background ranges from 55% to 90%, while it is always above 90% for a similar student with high parental background. Among students with high parental background, there is sizable variation in attainment across schools only for those with low innate ability. These results suggest that measures of school value added with respect to performance may convey most of the relevant information in the case of advantaged students, while multiple dimensions and more personalized assessments are necessary in order to understand how schools contribute to the educational outcomes of disadvantaged students.

Additional simulations show that retention policies are an important channel through which schools affect attainment. In fact, repeating a grade raises cognitive skills, but it steeply decreases preferences for schooling, and more so for students with a higher skill level. Overall, retention has a net strong adverse effect on the choice to pursue further education. Finally, I simulate a counterfactual policy that makes school attendance compulsory until the last grade of lower secondary education, rather than until turning 16 years old. Overall, the graduation rate would improve by 10 p.p., with an even larger increases (from 70% to 88%) among students with lower parental background. In fact, most students who drop out would have passed the final exams if they had stayed in school. These results, together with the findings on retention, indicate that policymakers interested in improving attainment in the population should give greater consideration to the interaction between existing rules and student decision-making.

1 Introduction

Higher educational attainment is associated with better labor market outcomes, and improved health and life satisfaction.[1] Furthermore, socioeconomic background is often the main determinant of educational attainment. Therefore, providing inclusive and high quality education that improves outcomes for all students, particularly the most disadvantaged, is a primary goal for policy makers all around the world. To address this issue, many countries have adopted school accountability measures for the purpose of monitoring school quality, deciding on corrective actions, and in some cases assigning funding. A prime example is the "No Child Left Behind" program in the US. In practice, school quality is often measured using nationwide tests, under the assumption that a school's ability to improve students' test scores is a sufficient measure of ability to improve their overall educational outcomes.[2]

I challenge this assumption by showing that school rankings based on performance are not informative of the overall impact that schools have on educational outcomes. In fact, I show that although school rankings based on standardized assessments are fairly successful in measuring a school's value added in terms of cognitive skills, they fail to capture a school's contribution to educational attainment, particularly for students from less favorable socioeconomic backgrounds. More specifically, I examine how the school in which a student is enrolled affects their performance, graduation probability, and probability of acquiring further academic education. To do so, I exploit administrative data on public middle schools in Barcelona (Spain), a setting in which 17% of students do not complete basic education and 35% do not enroll in high school. I find large variation across schools in their effect not only on cognitive skills development, but also on students' educational choices. Furthermore, given the limited correlation between school inputs in the different dimensions, attending a school with high value added in terms of performance does not necessarily increase the likelihood of pursuing further education. This is particularly relevant for subgroups of the population that are traditionally less likely to achieve high educational attainment.

The existing literature demonstrates that success in life is determined by more than just cognitive skills. Interventions aimed at improving a broader set of skills show impressive long-term returns and contribute to closing gaps due to socioeconomic background.[3] These results underscore the importance of moving beyond test scores alone in the debate surrounding school quality. Children are left behind not only when they receive low scores on a standardized test, but also when their school environment fails to develop both their cognitive and non-cognitive skills sufficiently and motivate them to pursue further studies. In this respect, secondary education is a crucial stage,

[1]For the OECD as a whole, the employment rate is 85% for tertiary-educated adults, 76% for adults who have completed high school, and less than 60% for those who have not. Moreover, adults of working age with a tertiary degree earn 54% more than those with only an upper secondary education, while those with less than a upper secondary education earn 22% less (OECD, 2018). Those with high literacy skills and a high level of education are 55% 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, as compared to 83% with lower educational attainment (OECD, 2016).

[2]The No Child Left Behind Act of 2021 (replaced by the Every Student Succeeds Act in 2015) requires public schools to administer an annual statewide standardized test. If a school repeatedly shows poor results, then various steps are taken to improve its performance. Since 1992, the UK has published the so-called "school league tables" which summarize the average GCSE results of state-funded secondary schools. Underperforming schools face various types of sanctions (Leckie and Goldstein, 2017).

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

because for the first time students can decide whether they want to continue their education and, at least to some extent, what they want to study. Thus, while in most countries basic education is compulsory, students can legally leave school when they reach a certain age (not necessarily upon completing a particular grade). Moreover, in many European countries, upon completing lower secondary education, students choose whether to enroll in the "academic track" which will provide them with access to university.[4] Students at risk of dropping out may benefit more from attending a lower-ranked school where they feel comfortable and are able to earn a diploma than a higher-ranked one from which they are likely to drop out. Similarly, the decision to continue into upper secondary or tertiary education may depend on previous performance, but also on non-cognitive factors such as a student's motivation or family support. The school environment may play an important role, substantially affecting a student's desire to acquire further education.

I develop and estimate a dynamic model of cognitive skills accumulation and schooling decisions throughout lower secondary education of students enrolled in heterogeneous schools. In each period, students acquire new skills and decide whether to continue their education or drop out. Upon graduation, they choose whether to enroll in academic upper secondary education, which consists of two additional years of high school that provide access to University. Importantly, before the decision of whether to drop out, a student can be required to repeat a year in order to stay in school, referred to as "retention". Retention may raise skills in the following period, but it also increases the time required to graduate and can therefore alter a student's preferences. Skills growth depends on innate ability, individual characteristics (particularly parental background), and school environment. Students have imperfect information about their level of cognitive skills, because they don't know their innate ability, though they progressively learn about it through the evaluations that they receive in school, thus performing a Bayesian updating. Their utility from schooling depends on their perceived cognitive skills, individual characteristics, and school environment. Additionally, the model allows the flow utility to vary with retention status.

The school environment is modeled through a vector of peer characteristics and a vector of school inputs. School inputs are parameters that capture school heterogeneity along multiple dimensions. First, schools differ in the way they contribute to the accumulation of cognitive skills for given peer quality. Second, they have different grading and retention policies, such that they vary in the probability of retaining students with given cognitive skills and individual characteristics. Third, they directly influence students' educational choices in various ways. The primary advantage of this structural approach is that it facilitates disentangling these channels based on the sequence of student's decisions and test scores. Another advantage is that it makes possible to quantify the relevance of informational frictions about one's ability in explaining educational choices. This may be important in explaining the drop out decision, particularly among retained students who often receive more negative signals.

The analysis is based on administrative data for the universe of students attending public middle schools in Barcelona during the period 2009-2015. As common in many European countries, nationwide tests are administered at the end of primary school and at the end of lower secondary education, but in the latter case several students have dropped out before taking the test. Moreover,

[4] For instance, students in Italy choose what high school to attend after completing grade 8, in France after completing grade 9, and in Spain after completing grade 10.

given the compulsory education laws in Spain, all students spend at least some time in lower secondary education and are evaluated by their teachers at least once, although grading policies vary across schools. The structure of the model makes it possible to combine the signals provided by these different evaluations, even if they are not directly comparable across schools, and to account for the self-selection of students who take the final standardized test.

The model is estimated using an approach that builds on James (2011) and Arcidiacono, Aucejo, Maurel, and Ransom (2016). I first estimate the production function of cognitive skills (in particular, the grade equations and the variance of unobserved ability), and individual beliefs over time using an Expectation-Maximization algorithm for computational convenience. I then estimate logit equations for the retention events. Finally, I estimate the parameters that govern the sequence of students' choices using maximum likelihood.

The results show that individual ability, parental background, and school environment are all important determinants of cognitive skills. Ability is found to explain around half of the variance in performance observed in the data. When compared to having parents with only basic education, having parents with tertiary education is associated with an increase in skills of up to one standard deviation. School environment has relevant impact by way of both peer quality and school inputs. For instance, the interquantile range of school inputs on skills is about 0.4 s.d. In other words, the effect on skill growth of attending a school in the top quartile rather than in the bottom one is comparable to that of having a highly educated father rather than a low-educated one.

Other things being equal, better parental education somewhat decreases the probability of retention and somewhat increases the probability of enrolling in high school. However, most of the effect of parental background on student attainment is due to its large contribution to cognitive skill levels. Conversely, schools have a sizable direct effect on educational choices beyond their indirect effect by way of cognitive skills. For instance, being in a school in the 75th percentile with respect to the distribution of the relevant school input rather than in the 25th percentile increases the utility from continuing one's education as much as an increase in cognitive skills of 0.5 s.d. Schools also vary widely in their retention policies: an increase in leniency of 1 s.d. reduces the probability of retention by as much as improving skills by 0.3 s.d. This is particularly important in view of the estimation results showing that retention negatively affects utility, particularly in the case of students with higher skill level.

Using the estimated parameters of the model, I simulate the educational outcomes that each student in the sample would have in each of Barcelona's school. Then, I compute three school rankings at the individual level: expected skills at the end of lower secondary education, probability of graduation, and probability of enrolling in academic upper secondary education, and compare them to aggregate school rankings based on the observational data, i.e. average test scores, proportion of graduates, and proportion of students enrolled in academic upper secondary education. The results indicate that the simulated school rankings by skill level are quite similar across students and highly correlated with the aggregate measure based on the observational data. In other words, a simple ranking based on average test scores is a good indicator of how the school contributes to a student's cognitive skills development, provided they do not drop put before completion. Conversely, individual school rankings by attainment (i.e. graduation and enrollment in further education) vary across students, are less correlated with the aggregate measures, and are generally different from the individual rankings based on skills. This implies that, on the one hand, the commonly used school ranking based on test scores is not informative about a student's educational attainment prospects. On the other hand, other simple statistics at the school level (such as the graduation rate) do not contribute much, because they do not accurately reflect the outcomes that students would have achieved if enrolled there, especially if their individual characteristics differ from the average of students in the school.

Differences in attainment across schools are sizable for the large majority of students with low parental education. For instance, the graduation probability of a representative student with average ability and low-educated parents ranges from 55% to 90%, while it is always above 90% for a similar student with highly educated parents. Among students with highly educated parents, there is sizable variation in attainment across schools only for those with low innate ability. These results suggest that measures of school value added with respect to performance may convey most of the relevant information in the case of advantaged students, while multiple dimensions are necessary in order to understand how schools contribute to the educational outcomes of disadvantaged students.

Additional simulations are used to further explore the role of retention and beliefs about ability. Although repeating a level raises cognitive skills, it has a net strong adverse effect on the choice to pursue further education. Schools vary widely with respect to retention policy, therefore this channel is crucial for students' attainment. Furthermore, retained students have somewhat lower beliefs about their ability than otherwise identical students who continue on the next grade. However, a counterfactual simulation without uncertainty confirms that most of the differences in their choices are due to changes in preferences, rather than misperception of their ability.

Finally, I simulate a counterfactual policy that makes school attendance compulsory until the last grade of lower secondary education, rather than until turning 16 years old. Overall, the graduation rate would improve by 10 p.p., with an even larger increases (from 70% to 88%) among students with low-educated parents. In fact, most students who drop out would have passed the final exams if they had stayed in school. Enrollment in high school would increase by 4 p.p. (6 p.p. among students with low-educated parents). The average skill levels of high school students, on the other hand, would be hardly affected, suggesting that future negative effects due to changes in peer composition are unlikely. These results, together with the findings on retention, indicate that policy makers interested in improving attainment in the population should give greater consideration to the interaction between existing rules and student decision-making.

Related Literature

The paper relates to several strands of the literature, particularly school quality, human capital development, and decision making. School accountability requires reliable measures of quality of individual schools (Allen and Burgess, 2013; Angrist, Hull, Pathak, and Walters, 2017; Kane and Staiger, 2002), or of school types (e.g. charter *vs* traditional public schools, as in Dobbie and Fryer (2011) or Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011)). This is typically accomplished using a value added approach, i.e. estimating the net effect of attending a given institution on a relevant outcome.[5] Test scores have been the most widely used outcome to capture

[5] For a survey, see Angrist et al. (2022). Previous studies also investigated the determinants of school value added. For instance Dobbie and Fryer (2013) and Angrist, Pathak, and Walters (2013) identifies practices such as increased

quality, under the assumptions that performance in school measures cognitive skills and that they are positively correlated with desirable outcomes in subsequent educational stages and in the labor market. The current analysis uses a comparable value added approach, but it explicitly models school inputs along multiple dimensions and studies the effect of attending a school with that combination of inputs on several outcomes. The results confirm the importance of moving beyond performance and rankings to study school quality.

Other strands of the literature have shown that cognitive skills alone do not explain educational choices and attainment that matter for future life outcomes. On the one hand, the literature on human capital development has established that the return on non-cognitive skills is no lower than that on cognitive skills and that the former may not be captured well by using test scores (for one of the seminal works, see Heckman and Rubinstein (2001)). On the other hand, the literature on decision making in Education shows that educational choices vary substantially among individuals with identical prior performances, for a multitude of reasons, ranging from differences in beliefs with respect to the return on each choice to differences in consumption value.[6] These studies focus primarily on individual traits and preferences, and on differences by gender and socioeconomic background.[7] The current analysis contributes to incorporate their findings within the research on the effect of school quality. Indeed, it seems plausible that the school environment, like the home environment, may substantially contribute to non-cognitive skill development, taste formation, and provision of information regarding the returns on education.[8] A structural model of cognitive skills development, retention, and choices makes it possible to first estimate the effect of school environment on cognitive skills, and then to estimate its direct effect on a given educational choice.

The current analysis differs from recent contributions (Angrist, Cohodes, Dynarski, Pathak, and Walters, 2016; Deming, Hastings, Kane, and Staiger, 2014) that study the effect of attending a given school on measures of attainment, such as high school graduation, college enrollment, and college persistence. They use those outcomes (rather than test scores) as alternative measures of human capital, to confirm that the gain found by previous studies is not due to a "teaching to the test" attitude, but rather an actual improvement in skills. Their analyses cannot assess

instructional time, high-dosage tutoring, and high expectations which are responsible for the success of some kind of charter schools, while Fryer (2014) shows that some of these practices can be successfully exported to public schools.

[6]Several papers have looked at the choice of college major in the US (for a survey, see Altonji, Arcidiacono, and Maurel (2015)). Avery, Hoxby, Jackson, Burek, Pope, and Raman (2006) and Hoxby and Avery (2012) study the role of financial constraints and information in the applications to selective colleges of high-achieving low-income students. 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 the major chosen, heterogeneous tastes are the dominant factor. On the other hand, Wiswall and Zafar (2015) find that college students are misinformed to a great exxtent about population earnings and revise their earnings beliefs in response to new information. Kinsler and Pavan (2015), Bordon and Fu (2015), Hastings, Neilson, and Zimmerman (2013) exploit Chilean data to study choice of college and major.

[7]Recent examples include Belfield, Boneva, Rauh, and Shaw (2020), who collect survey data on students' motives in pursuing upper secondary and tertiary education in the UK. They find that beliefs about future consumption values play a more important role than beliefs about monetary benefits and costs, and differences in the perceived consumption value across gender and socioeconomic groups can account for a sizable proportion of corresponding 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.

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

whether improvements in graduation rate or college enrollment are due only to the improvement in cognitive skills as measured by standardized tests or to other factors as well. The advantage of the current structural model is that it allows to disentangle effect of schools on attainment by way of cognitive skills and that by way of other channels, and to examine their interaction. The results are consistent with contemporaneous studies that show that a school's impact on high-stakes tests is weakly related to educational attainment and other important life outcomes, such as crime or teen motherhood (Beuermann, Jackson, Navarro-Sola, and Pardo, 2022), while a school's impact on non-cognitive skills can be important in explaining them (Jackson, Porter, Easton, Blanchard, and Kiguel, 2020). Their findings also suggest that one-dimensional indicators based on test scores are not sufficient in order to measure school quality.

I find that the school environment is particularly relevant to educational choices and attainment in the case of students with low-educated parents or low prior cognitive skills. Dearden, Micklewright, and Vignoles (2011) question the use of a single measure of value added on performance in order to assess school effectiveness by showing that a school's effect can be differ according to prior ability levels. The current analysis shows that even if the performance of every student is affected similarly by a given school environment, its effect on educational attainment may vary according to background and ability.

Finally, the current study contributes to the debate on the effectiveness of grade retention. Even though retention is a common practice in many countries, the empirical literature provides mixed evidence of its effectiveness in improving student performance (Allen, Chen, Willson, and Hughes, 2009; Fruehwirth, Navarro, and Takahashi, 2016). The results discussed here suggest that it can improve test scores at the end of middle school (at the cost of longer time spent in education); however, it has a negative effect on a student's consumption value of schooling. The net result is a large increase in drop out rates among retained students and a lower probability of enrollment in high school.[9] Interestingly, the gap is larger for students with a relatively high level of cognitive skills who would not be at risk of retention in more lenient schools.

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 summarizes the estimation procedure. Section 5 presents the estimation results. Section 6 discusses simulations and counterfactual analysis, with 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 attending one of 47 public middle school in Barcelona (Spain) in 2009 or in 2010. I exploit various data sources to collect detailed information on enrollment, school progression, performance, and sociodemographic characteristics. This section provides some background on the school system, describes the data sources, and discusses the descriptive statistics.

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

2.1 The education system

Basic education in Spain 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, students spend 6 years in primary education, followed by 4 years in middle school. All students begin primary school in the year in which they turn 6 years old. Retention during primary education is uncommon and, by law, it can occur no more than once. Thus, students usually start middle school in the year in which they turn 12, or at most one year later. In middle school, students can be retained at most twice, but not in the same grade. While repeating a grade during lower secondary education is fairly common, repeating twice is rare. About 20% of students in public schools in Barcelona are retained once, but only 3% are retained twice. Thus, students can graduate in the year in which they turn 16 or later, depending on their retention history.[10]

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

After successfully completing lower secondary education, students can enroll in high school for two more years (corresponding to ISCED level 34, general upper secondary education). Alternatively, they can also choose vocational training, but it does not provide 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, limits on class size, and teacher assignment.[11] On the other hand, schools have extensive autonomy in deciding how to evaluate student performance and whether to advance them to the next grade.

Families have relatively limited options when choosing a middle school for their children. In fact, every primary school is affiliated with one or more middle schools and students from an affiliated primary school have priority if the middle school is oversubscribed (other priority criteria such as distance between school and home are also used, if necessary). The structure of the application process creates a major incentive to designate the school for which the student has the highest priority as first choice, because students who are not admitted to their first choice lose their priority for other schools. For instance, 92% of families in Barcelona applied to an affiliated middle school in 2009 and 88% to the closest middle school.

While lower and upper secondary education are two separate stages, they are usually located in the same school ("Instituto de Educación Secundaria"). Thus, the principal is the same for both and classes take place on the same premises. Typically, there are different teachers in each, although some might switch between them over time. All public schools offer two main tracks in upper secondary education, namely Science and Humanities, and admittance is guaranteed to all students who graduate middle school. Arts, is offered as a third track in a small number of schools

[10]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)

[11]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, all students study core subjects such as Mathematics, Spanish, Catalan, and English. They can also chose to attend a limited number of elective subjects, whose evaluations are not part of this study. Further details on the allocation of students to classes are provided in Appendix E.

but there is no guaranteed admission in this case. About 4% of the students in Barcelona choose Arts, 43% choose Humanities, and 53% choose Science. About 95% of the students in Science and Humanities stay in the same school for both upper and lower secondary education.

2.2 Variables

The panel used for the analysis consists of 5140 students in 47 middle schools in Barcelona who began their lower secondary education in either September 2009 or September 2010. They are observed from their last year of primary education (2008/2009 or 2009/2010) to their last year of lower secondary education (up to 6 years later). Moreover, for those who successfully graduate, it is observed whether or not they enroll in academic upper secondary education subsequently. Data on school progression and performance and data on socioeconomic background are taken from several administrative data sources, including the Catalan Ministry of Education and the national census. Data sources are described in more detail in Appendix B.1, while the data cleaning and the various steps in arriving at the final sample are discussed in Appendix B.2. This subsection introduces the variables used in the empirical analysis.

2.2.1 Sociodemographic characteristics

Demographic characteristics. The data includes dummies for gender and immigrant, and two variables for the exact age of the student. More specifically, the dummy "retained in primary school" takes a value of 1 if the child turns 13 rather than 12 in the first year of middle school. The other variable is the day of birth, rescaled to the interval 0-1, such that 1 is January 1st and 0 is December 31st. Furthemore, students in the sample belong to one of two cohorts: those who began lower secondary education in 2009, and those who began in 2010.

Parental education. Mother's and father's education take three values: Low (at most lower secondary education), Average (upper secondary education), and High (tertiary education). Two dummies are included in order to capture missing information for the mother or the father. In some cases, I use three categories of parental background, which are created by averaging the mother's and father's education: "Low" if both parents have at most lower secondary education, "High" if one of them has tertiary education and the other has at least upper secondary education, and "Average" in the remaining cases.[12]

Application to middle school. The dummy "school not first choice" takes a value of 1 if the middle school that the student attends was not their first choice.

2.2.2 Neighborhood quality.

I create an index of neighborhood quality based on average gross income and on proportion of highly educated mothers and fathers in the postal code of the student's home address.[13] This

[12] If information is missing for either parent, then only the education level of the other is used.

[13] About 96% of the students in the sample live in Barcelona, which is divided into 42 postal code areas. The remaining 4% come from nearby municipalities. Average income at the postal code level is available for large municipalities (Badalona, Hospitalet de Llobregat, and Sabadell) and at the town level for the few smaller municipalities

index is the first component of a principal component analysis. It explains 92% of the variance and loads similarly on the three factors.[14] The index is normalized to have mean 0 and standard deviation 1 in the sample.

2.2.3 Performance and retention

External evaluations. Students undergo an externally administered region-wide assessment in Math, Catalan and Spanish at the end of primary education and at the end of lower secondary education. Outcomes are normalized to have mean 0 and standard deviation 1.

Internal evaluations. End-of-the-year evaluations are carried out by teachers in each subject. The final grade takes in account the results of several tests administered during the year. For comparability with the region-wide exams, I use the average of the grades in Math, Spanish, and Catalan, normalized to have mean 0 and standard deviation 1.

Retention. At the end of the year, students either successfully advance to the next grade or are retained, namely they have to repeat the same grade if they wish to stay in school. In the analysis, dummies for end-of-the-year retention are aggregated in various ways in order to define variables over a longer time period, such as, for example, being retained before the school year in which the student turns 16.

2.2.4 School variables

The empirical analysis uses a vector of dummies to identify the school in which a student is enrolled. The data makes it possible to identify a student's class in a given year, and therefore peer variables can be computed at the class level. More specifically, peer variables are defined as the average value among all the peers in the class, including students who do not belong to the sample (because, for example, they were retained and belong to an older cohort).[15]

Two peer variables are used in the main analysis: the proportion of females and an index of peer quality. The latter combines information on the proportion of college graduates among the students' mothers and fathers, the proportion of immigrant students, the average test score at the end of primary education, and the share of students who took that test.[16] The index is the first

that appear in the sample.

To the best of my knowledge, information on the education of residents is not available at the postal code level. Instead, I use enrollment data in Catalonia to compute the proportions of college educated mothers and fathers among the parents of children aged 6 to 15 living in each postal code area in the relevant years. While these measures are only a proxy for the proportion of college graduates in the overall population, they refer to the group of adults that is likely to be more relevant for teens enrolled in school.

[14] I replicated the analysis introducing the three variables separately. The results are very similar and the coefficients of the other regressors remain virtually unchanged. Therefore, I preferred the more parsimonious specification with only one index.

[15] Defining peers at the class level rather than at the school level appears to be preferable because in the Spanish system students in the same class are exposed to the same teachers and the same contents, and are always together in school. Moreover, this allows peer variables to vary both over time and within school. Given the limited number of cohorts this is a desirable feature. Class allocation is further discussed in Appendix E.

[16] The proportion of students for whom I could not retrieve test score is also a proxy for recent immigrants who may have limited knowledge of the local languages and therefore may have been exempted from the test or may have completed primary education in another country.

component of a pca and explains 73% of the variance. The loadings are very similar in magnitude ranging from 0.42 to 0.47. The proportion of immigrant peers has a negative sign while the others are positive.[17] Finally, the two peer variables are standardized to have mean 0 and standard deviation 1 in the sample.[18]

To estimate the model of educational choices, I also need to define counterfactual peers for events that did not take place, such as, for example, the peers that a retained student who dropped out would have had if they had not dropped out. For this purpose, I use the average peer variables for their classmates who were retained but did not drop out.[19]

2.3 Descriptive statistics

About 17% of the students in the sample do not graduate middle school. Most of them leave school voluntarily before completing lower secondary education: 8.6% drop out as soon as possible, while the rest remain for one or more additional years after reaching the legal age to drop out. About 66% of the initial pool of students eventually enroll in high school (78% of those who graduated).

Table 1: Descriptive statistics by subgroups of the population

			Test score		Drop out	Attainment		Peer	Neighbor.	School
	N	%	at entry	final	at 16	grad.	high sch.	quality	quality	1st 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. "Test score at entry" (final) measures results in a region-wide test administered at the end of primary school (middle school), "School 1st choice" takes value 1 if the student is enrolled in the preferred school.

As shown in Table 1, there is significant variation in the descriptive statistics 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 drop out at 16, and only 70% of them complete lower secondary education (as compared to 95% of children with highly educated parents). Moreover, those who graduate have on average lower test scores (-0.38 s.d. versus 0.59 s.d. for students with highly educated parents) and are less likely to pursue further studies. Only 43% of the initial pool

[17] The correlation between peer quality and proportion of females is virtually 0. When the proportion of females is included in the pca and two (rotated) components are extracted, the second component is almost identical to the proportion of female.

[18] I replicated the analysis using each of the peer regressors separately. Overall, the results are fully aligned although some parameters are not precisely estimated. Therefore, I prefer the more parsimonious specification.

[19] In most cases, students stay in the same class throughout middle school, except for those who are retained or drop out. Moreover, a retained student may be assigned to another class, for example if it has more free places. Thus, counterfactual peer variables are a weighted average of the possible groups of peers that a retained student would have had in the counterfactual scenario. In the very few cases in which there are no peers in the relevant situation, such as for example if no one in the class repeated the year, I use peers at the school level.

of students with low-educated parents enroll in high school, as compared with 86% of students with highly educated parents.

Importantly, Figure 1 shows that educational outcomes vary even among students who have similar performance at the beginning of lower secondary education, as measured by the standardized assessment at the end of primary education ("test score at entry" in the figure). Panel (a) plots the average graduation rate for each decile of test score at entry for students with low-educated parents (blue dots) and those with highly educated parents (red dots). The latter have much higher graduation rate in every decile. For example, the graduation rate in the second decile is already 90% for students with highly educated parents, as opposed to only 70% for students with low-educated parents. Similarly, panel (b) shows that students with highly educated parents have substantially higher odds of continuing in to high school in each decile and, if anything, differences are larger among students whose test score at entry is above average. The other panels suggest that there are several channels contributing to these differences. Students with low-educated parents are more likely to fail a grade for any level of test score at entry (panel (c)). For example, more than 20% of students with above average scores at the end of primary education repeat a grade, while the proportion is negligible for those with highly educated parents. Moreover, their test scores diverge during lower secondary education: students with low-educated parents do worse according to both internal and external evaluations (panels (d) and (e), respectively). Interestingly, they are also less likely to enroll in high school after graduation in every decile of final test score besides the first one (panel (f)).

According to Table 1, there is also a large gap between students from an immigrant background and natives: the former show lower performance and have lower probabilities of completing middle school and enrolling in high school. Boys and girls have on average similar performance both at the beginning and at the end of middle school. However, there are salient differences in their educational attainment: boys are more likely to drop out as soon as possible, 20% of them do not complete middle school (as compared to 14% of girls), and only 60% enroll in high school (as compared to 71% of girls).

About 93% of students attend the middle school that was the first choice in the application list. The rate is very high among all the subgroups.

Disadvantaged students are more likely to have classmates from a similar background and to live in a neighborhood with a lower socioeconomic status. The index of peer quality is about 0.5 s.d. less for students with low-educated parents, and the index of neighborhood quality is lower by about 0.4 s.d. Meanwhile, boys and girls have similar peers and live in similar neighborhoods.

The last four rows of Table 1 give the descriptive statistics by retention status. Students are grouped into four categories: those who were never retained before leaving middle school; those who were retained in primary school (8% of the sample); those who were retained for the first time in middle school before reaching the last grade (15%); those who were retained for the first time in the last grade (4%). Students who were already behind before turning 16 are more likely to drop out early, especially if they were retained during secondary education: 30% of them immediately drop out, while only 3% of non-retained students do so. Moreover, less than half of them graduate middle school and very few enroll in high school. Students who repeat the last grade are less likely to graduate and enroll in high school althoug they have better odds than students retained at an earlier stage.

Table 2 describes the distribution of incoming students' characteristics and their outcomes by school percentile and shows that schools are quite different both in the types of students they teach and in the outcomes they produce. For instance, for a school at the 75th percentile, incoming students have primary school test scores that are on average 0.5 s.d. better than those in the school at the 25th percentile. Moreover, the interquantile range (IQR) of the proportion of students with highly educated parents is 30 percentage points. The IQR of the final test score is almost 0.6 s.d. and that for enrollment in high school is 21 p.p. On the other hand, in all schools only a small share of students had a different school as their first choice (even the school at the 25th percentile was the first choice of 90% of its students).

Table 2: Descriptive statistics by schools

	Highly edu. parents	Spanish	Test score at entry	Graduate	High school	Test score final	Neighbor. quality	School 1st 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.

Figure 2 provides further evidence of the large variation in outcomes across schools. It plots the average test score at the end of lower secondary education on the x-axis and the proportion of students that graduated (left) or the share of students that enroll in high school enrollees (right) on the y-axis. Performance and attainment at the school level appear to be positively correlated.[20] Figure 3 shows the same outcomes by parental background. It can be seen that variation in attainment is particularly large for students with low-educated parents. Moreover, within-group correlation between test scores and attainment appear to be lower than the overall correlation. This is confirmed in in Figure A-10 in the appendix, which shows that the correlation decreases considerably once parental education is controlled for. This descriptive evidence suggests that average final test score at the school level is a poor proxy for the educational attainment of students with low parental background.

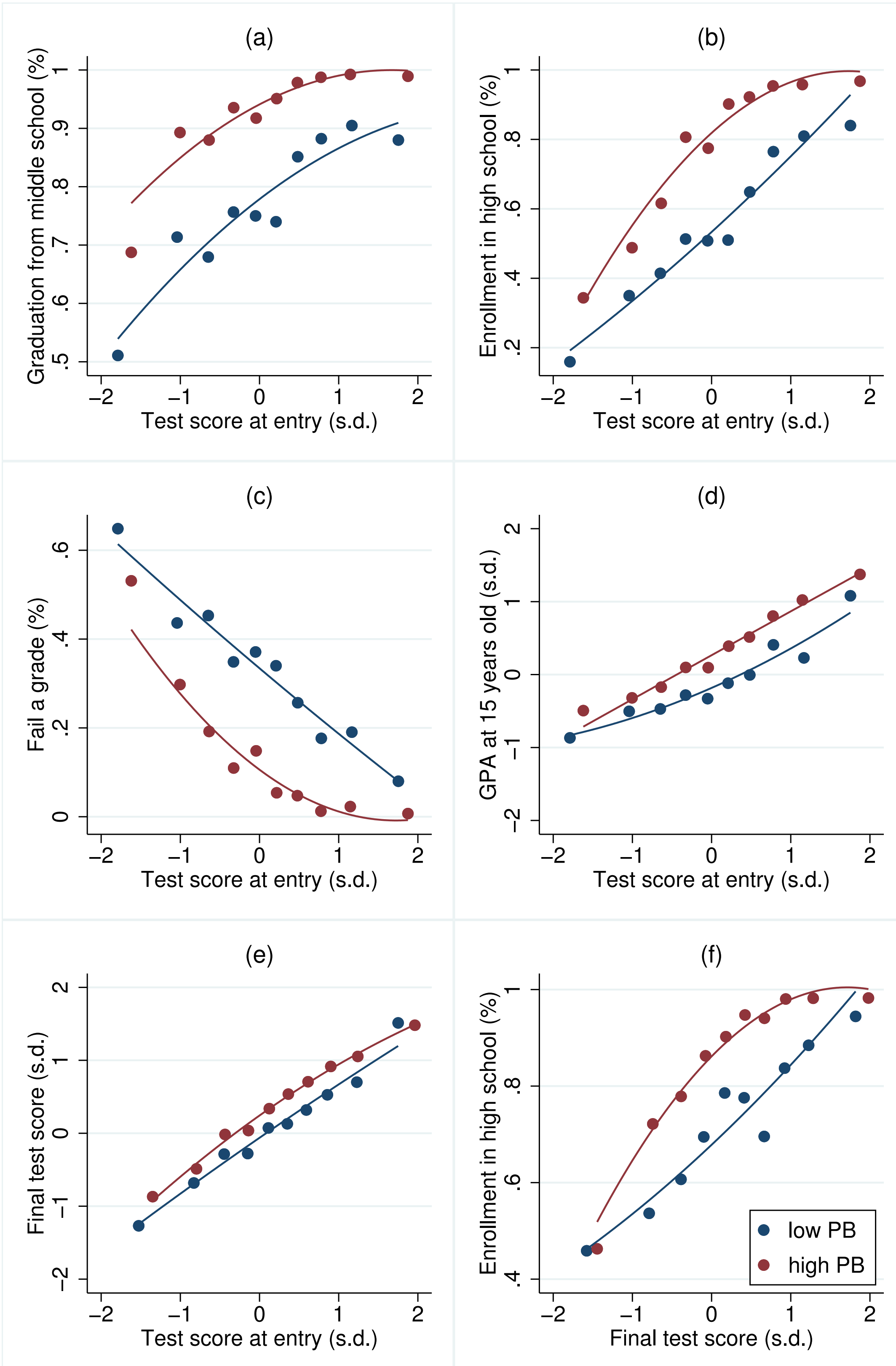
3 Model

3.1 Overview

A model is constructed to explain cognitive skills development and educational choices among students enrolled in secondary education. The latter include whether to stay in school after the legal age to drop out, and whether to pursue further academic education. While in school, students

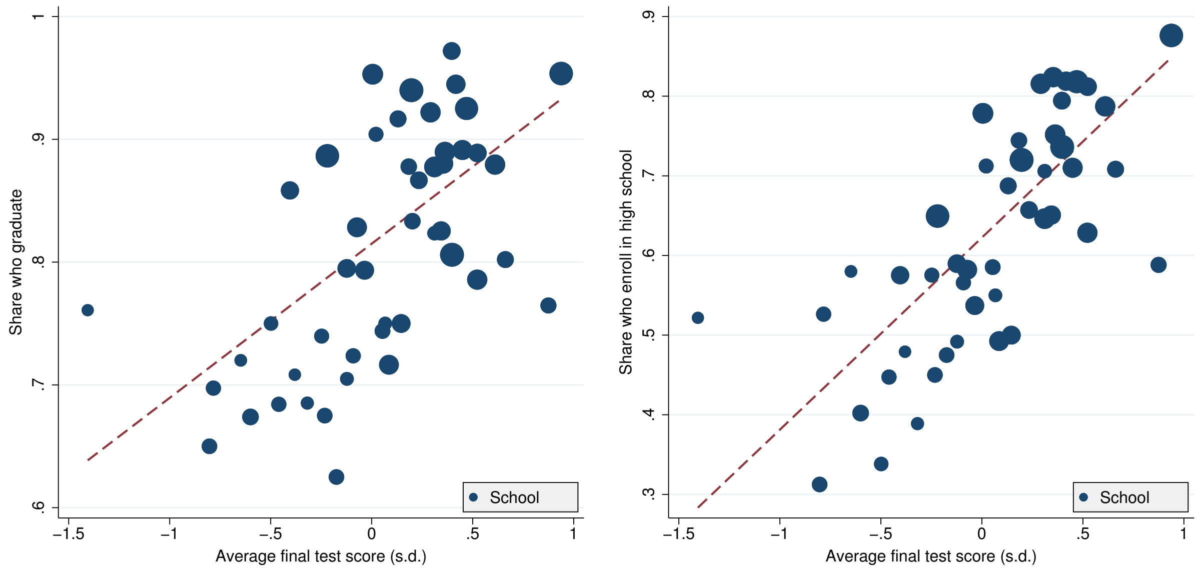
[20] The correlation in the left panel is 0.59, while that in the right panel is 0.72 . Weighted correlations by school size are 0.57 and 0.74, respectively.

Figure 1: Educational outcomes by parental background



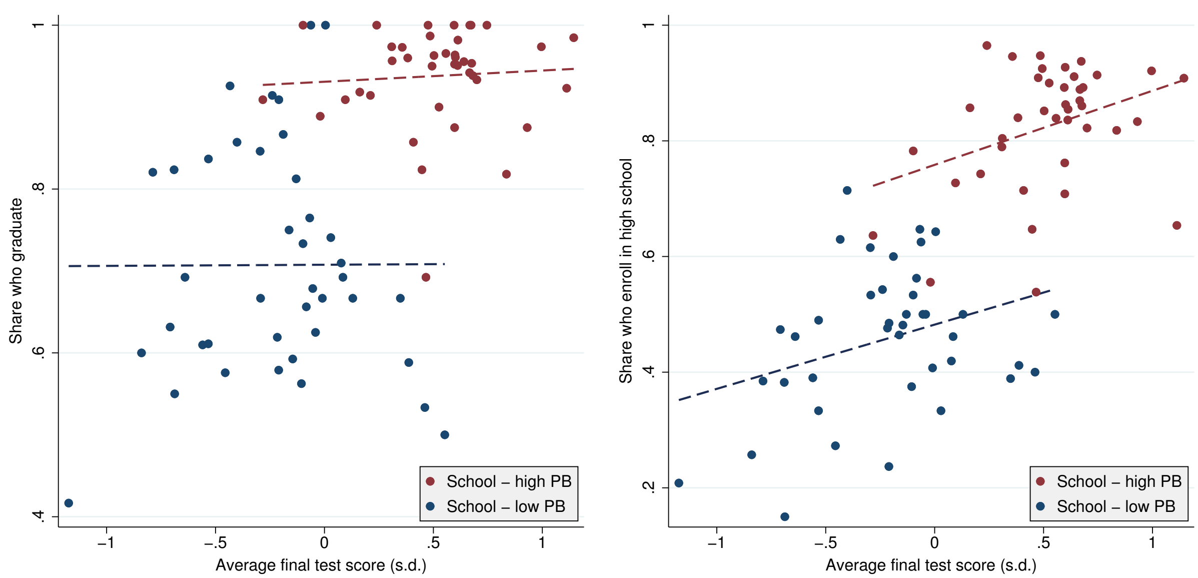
Note. Statistics in panels (a) to (d) are computed using all students in the sample. Statistics in panels (e) and (f) use the sub-sample of students who graduate from lower secondary education. In panel (a) to (e) the population is allocated in deciles by test score at the end of primary education; in panel (f) the deciles are computed using the test score in the last grade of lower secondary education. The dots plots the average value in each decile of the variable displayed in the y-axis by parental education. The continuous lines are quadratic fits of the data.

Figure 2: Educational outcomes by school



Note. Each dot plots average statistics at the school level. The dot size is proportional to the school size. The dashed line is a linear fit

Figure 3: Educational outcomes by school and parental background



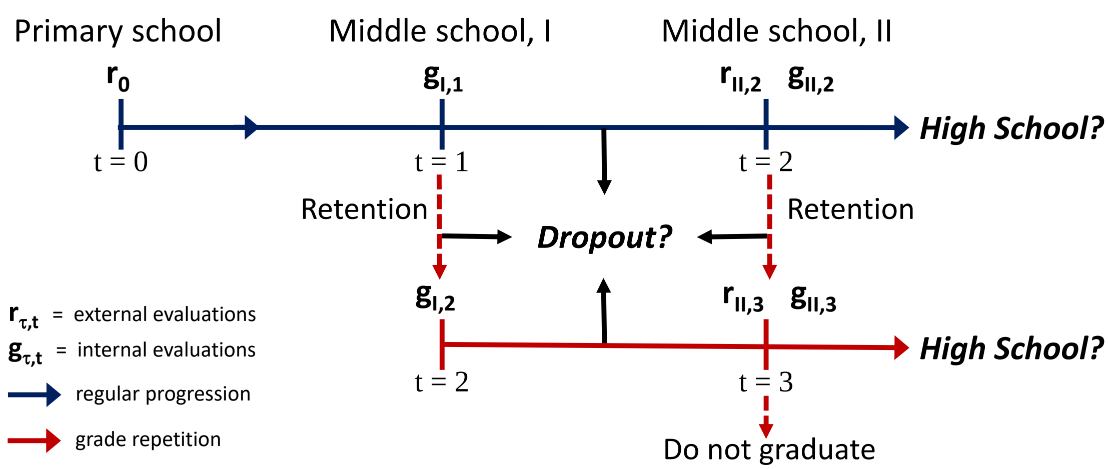
Note. Each dot plots average statistics at the school level for students with low parental background (blue dot) or high parental background (red dot). Dashed lines are linear fits.

might fail a grade and have to repeat it, which may affect their incentive to continue, especially because it prolongs the time required to graduate.

Cognitive skill accumulation depends on previous skill level and on contemporaneous inputs: individual characteristics (including ability h which is unknown) and school environment. While in school, students are evaluated and use the results to infer their level of ability, and corresponding cognitive skills. There are two type of evaluations: 1. standardized test, which are the same across schools; 2. internal grades, which have school-specific components.

Retention is probabilistic and depends on a student's cognitive skills, individual characteristics and school environment. Students are assumed to be forward-looking and make choices that yield the highest expected utility. Their flow utility at each point in time depends on what they believe to be their cognitive skills, on their individual characteristics, on the school environment, and on their retention history.

Figure 4: Timeline of the model



The model is constructed to reflect the Spanish education system with some unavoidable simplifications. The timeline is summarized in Figure 4.

3.1.1 A student's progression in school

Timeline. At time $t = 0$, student i is evaluated by means of a nation wide test ($r_{0,i}$), completes primary school, and begin middle school s. Lower secondary education covers two stages: level I and level II. Each level takes one time period, but students may be retained once, during either level I or level II. In that case, they have to spend an additional period in the same level, unless they drop out.[21]

At time $t = 1$, students finish the first period in school, undergo an internal evaluation ($g_{\mathrm{I},i1}$) and are informed as to whether they can continue on to level II ($\mathrm{fail}_{\mathrm{I},i1}$). Education is no longer compulsory after $t = 1$, and therefore students decide whether to stay in school or drop out:

- Retained students who don't drop out repeat level I. At time $t = 2$, they again undergo an internal evaluations ($g_{\mathrm{I},i2}$) and can advance with certainty to level II.

[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 of lower secondary education.

- Not-retained students who don't drop out continue to level II. At time $t = 2$, they undergo an internal evaluation and a nationwide external evaluation ($g_{\text{II},i2}$ and $r_{\text{II},i2}$). They are also informed as to whether they have successfully completed lower secondary education or have been retained ($\text{fail}_{\text{II},i2}$).

Thus, at time $t = 2$, students who have not successfully completed lower secondary education decide whether to drop out or to stay in level II. If they do not drop out, then at time $t = 3$ they undergo internal and external evaluations ($g_{\text{II},i3}$ and $r_{\text{II},i3}$), and are informed as to whether or not they are graduating ($\text{fail}_{\text{II},i3}$). If they do not graduate at $t = 3$, they must leave school without obtaining a diploma.

Students who successfully complete lower secondary education – whether at $t = 2$ or at $t = 3$ – decide whether to continue on to high school.

Observable characteristics. Student characteristics x_i, which directly affects skills, retention, and choices, include gender, citizenship, parental education, year and day of birth, cohort, an index of neighborhood quality, and a dummy for not attending one's firts-choice school (as described in Section 2.2).

Unobserved ability. Cognitive ability h is an innate trait, randomly distributed in the population according to a normal distribution $\mathcal{N}(0, \sigma)$. A student's cognitive ability h_i is uncorrelated with their individual characteristics x_i. Students do not know their level of ability, but they know the distribution in the population. When they receive new (normally distributed) signals about their level of ability, they update their beliefs and formulate a posterior distribution. Ability is one of the determinants of cognitive skills.

3.1.2 The role of schools

The school environment affects a student's cognitive skills, progression and choices through a vector of peer characteristics p_{it} and a vector of school inputs $\mathcal{S}$. As described in Section 2.2, when bringing the model to the data, two variables capture peer characteristics: an index of peer quality, based on classmates' previous performance and family background, and the proportion of females. To account for non-linearity in peer effects, a third-degree polynomial in each variable is used. Furthermore, each school is characterized by a vector of four school inputs $\mathcal{S} = (\mathcal{A}, \mathcal{I}, \mathcal{M}_M, \mathcal{M}_A)$, which are estimated using the model and can be interpreted as measures of school value added, above and beyond peer quality in that school. To simplify the notation, the subscripts i and s are omitted from the school inputs, such that they always refer to the school in which the student is enrolled.

$\mathcal{A}$ is the school's contribution to cognitive skill development during lower secondary education (as described in Section 3.2). $\mathcal{I}$ is the school grading policy, or in other words the school's level of leniency in grading and deciding whether to retain. It has a direct effect on internal evaluations (see 3.2.3), retention and graduation (see 3.3), and the flow utility from school attendance (see 3.4.1). While the model does not explicitly incorporate a student's effort, the direct effect of $\mathcal{I}$ on utility captures the fact that students may have higher utility in schools in which they need to invest less effort in order to advance to the next grade or achieve a given evaluation score. $\mathcal{M}_M$

and $\mathcal{M}_A$ are the school's inputs that affect the choice of whether to dropout of middle school and whether to continue on to high school, respectively (see 3.4.1). In other words, they are the school's contribution to "tastes" or "motivation" for education, beyond the school's effect on skills and the grading policy in place. Given that both $\mathcal{I}$ and $\mathcal{M}_M$ ($\mathcal{M}_A$) have a direct effect on the utility from education, it is convenient to also define their joint effect $\mathcal{T}_M$ ($\mathcal{T}_A$).

To illustrate, schools with high teacher quality can be thought of as having high $\mathcal{A}$, while schools with lenient grading standards can be thought of as having high $\mathcal{I}$. Furthermore, more inclusive schools that provide students at risk with psychological support to avoid dropping out may exhibit high $\mathcal{M}_M$. Schools that organize orientation events and emphasize the benefits of pursuing further studies may exhibit high $\mathcal{M}_A$.

Importantly, the total contribution of the school environment on the observed outcomes depends on the interplay between its various components. For instance, the school environment directly affects the choice of enrolling in upper secondary education through its effect on tastes, and indirectly, because it affects cognitive skill development (higher skill levels may increase the utility from education) and the student's progression through lower secondary education (by way of retention policy and the choice whether to drop out).

3.2 Cognitive skills formation and evaluations

3.2.1 Skills

The creation of cognitive skills is a cumulative process. Their level in a given time period depends on the level previously achieved and on contemporaneous determinants. Student i starts secondary education with skills $C_{i,0}$, gains $C_{i,\text{I}}$ at the end of the first level, and $C_{i,\text{II}}$ at the end of the second level. When they repeat a level, the most recent attainment of skills replaces what was attained in the previous time period. $C_{i,\tau t}$ denotes the cognitive skills of student i in period τ at time t:

$$C_{i,0} = z_i' b_0 + m_0 h_i \tag{1}$$

$$C_{i,\tau t} = a_\tau C_{i,\tau-1} + z_{it}' b_\tau + m_\tau h_i, \qquad (\tau, t) \in \{(1,1),(1,2),(2,2),(2,3)\} \tag{2}$$

In each period, the contemporaneous determinants are ability $h_i \sim \mathcal{N}(0,\sigma)$ (which is unknown), and individual and school variables. At time $t = 0$, the observed determinants are the individual time-invariant characteristics x_i and primary school effects $\mathcal{P}_i$. Starting from $t = 1$, $z_{it}' b_\tau$ is a linear function of x_i, a dummy rep_{it} (which takes a value of 1 if the level is being repeated for the second time), peer variables p_{it}, school input $\mathcal{A}$, and interaction between school input and certain individual characteristics.[22] In practice, interactions between school input and gender, nationality and parental education dummies are included. This specification allows a school to have differential effects on students according to their characteristics while remaining parsimonious with respect the

[22] With some abuse of notation, I define $z_{it}' b_\tau$ for student i enrolled in s at time t in level τ as follow:

$$z_{it}' b_\tau = \sum_{j=1}^{J} b_{\tau,x_j} x_{ij} + \mathcal{A} + \sum_{j=1}^{J'} b_{\tau,sx_j} (\mathcal{A} x_{ij}) + b_{\tau,rep} rep_{it} + \sum_{k=1}^{K} b_{\tau,p_k} p_{itk}, \tag{3}$$

where J is the number of individual characteristics, J' is the number of interactions ($J' \leq J$), and K is the number of peer characteristics. For clarity, I always use $z_{it}' b_\tau$ or similarly defined terms in the remainder of this section.

the number of parameters to estimate.[23]

3.2.2 External evaluations

The nationwide test score at time t in level τ is an unbiased measure of cognitive skills, that is, an affine transformation plus an exogenous normally distributed error:

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

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

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

and those who stay in school long enough also receive:

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

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

3.2.3 Internal evaluations

At the end of each period in secondary education, students receive an evaluation from their teachers. 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, such that they may design and administer more or less difficult tests and are more or less lenient 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, and the school in which they are enrolled. Finally, as in the case of nationwide test scores, there is an exogenous normally distributed error:

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

Note that in principle all of the contemporary observed determinants of cognitive skills can be a source of discrepancy between internal and external evaluations. In particular, differences in grading policies across schools are captured by school inputs $\mathcal{I}$, such that the higher $\mathcal{I}$ is, the more lenient is the school. Conversely, unobserved ability h_i only affects evaluations through cognitive skills.

Finally, the error terms of internal and external evaluations may have different variances. This would be the case, for instance, if internal evaluations were more precise (even if "biased") because they average several tests administered during the year.

[23] A specification that estimates different school inputs according to a student's characteristics would be overly demanding in the current setting.

3.2.4 Signals and posterior distribution

I assume that students know the parameters that govern cognitive skills and grading, but do not observe h_i, and therefore they do not know $C_{i,\tau t}$ exactly at any point in time. Given that h_i enters the skill equation linearly and it is the only unobserved determinant, there is a one-to-one correspondence between beliefs about ability and beliefs about skills. Students infer signals on h_i from evaluations and subsequently update their beliefs about their level of cognitive skills in a Bayesian fashion. More specifically,

$$s(r_{\tau,it}) = h_i + \frac{1}{\lambda_\tau}\epsilon_{r_\tau,it} \tag{8}$$

$$s(g_{\tau,it}) = h_i + \frac{1}{\mu_\tau}\epsilon_{g_\tau,it}. \tag{9}$$

All students observe $r_{0,i}$ and $g_{\mathrm{I},it}$, while the other signals 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, the posterior distribution is updated using the previous posterior as prior.[24]

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

3.2.5 Final grade equations and parameters to estimate

The scale factors $\lambda_\tau, o_\tau, \mu_\tau, \nu_\tau$ 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, skill parameters in level I and level II are identified "up to scale", that is, meaningful comparisons of skill values can be done within a level, but not across levels. Finally, a necessary assumption for identification is that school and teachers' grading policy is constant across levels, i.e. $\gamma_{\mathrm{I}} = \gamma_{\mathrm{II}} = \gamma$.[25]

The final grade equations – which I bring to the data to estimate the parameters – are obtained by substituting the definition of cognitive skills (2) in the evaluation equations defined in (4) and

[24]See DeGroot (1970). For instance, suppose that a student with ability h is in level II and undergoes 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 a 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 $\Sigma = \begin{bmatrix} \omega & \omega & \omega \\ \omega & \omega + \rho_{r_{\mathrm{II}}} & \omega \\ \omega & \omega & \omega + \rho_{g_{\mathrm{II}}} \end{bmatrix} = \begin{bmatrix} \omega & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. Thus, the posterior distribution of h after receiving signals $s = \widehat{s}$ is simply the conditional distribution of h with normal distribution $\mathcal{N}(\overline{\mu}, \overline{\omega})$, where $\overline{\mu} = \mu + \Sigma_{12}\Sigma_{22}^{-1}(\widehat{s} - s)$ and $\overline{\omega} = \omega - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

[25]With regard to teachers' grading policy, the assumption is necessary because external evaluations are observed only in level II. the fact that parameters are identified "up to scale" is a consequence of the fact that evaluations are not fully comparable across levels. For instance, getting a 8/10 in first grade, require a different set of skills than getting the same score in the last grade. A student who scores a 8/10 in first grade, would probably fail a test in the last grade, because they haven't yet acquired the necessary knowledge.

(7), and redefining some of the coefficients:

$$r_{0,i} = z'_{i0}\beta_0 + h_i + \epsilon_{r_0,i} \tag{10}$$

$$g_{\mathrm{I},it} = \nu_{\mathrm{I}} + z'_{it}(\beta_{\mathrm{I}} + \gamma) + \kappa_{\mathrm{I}} I_{0,i} + \mu_{\mathrm{I}} h_i + \epsilon_{g_{\mathrm{I}},it} \tag{11}$$

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + z'_{it}\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i} + \lambda_{\mathrm{II}} h_i + \epsilon_{r_{\mathrm{II}},it} \tag{12}$$

$$g_{\mathrm{II},it} = \nu_{\mathrm{II}} + \mu(z'_{it}\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i} + \lambda_{\mathrm{II}} h_i) + z'_{it}\gamma + \epsilon_{g_{\mathrm{II}},it}, \tag{13}$$

where $I_{\tau-1,i}$ is the portion of previous cognitive skills that is due to 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. Furthermore, I define $\mu = \frac{\mu_{\mathrm{II}}}{\lambda_{\mathrm{II}}}$.

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

3.3 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 continue on to level II. At time $t = 2$, students who are in 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 school) or they graduate.

I assume that the conditional probability takes a logit form:

$$\Pr(\text{fail}_{\tau,it} = 1 | w_{it}) = \frac{\exp(w'_{it}\zeta_\tau)}{1 + \exp(w'_{it}\zeta_\tau)}, \qquad (\tau, t) \in \{(\mathrm{I}, 1), (\mathrm{II}, 2), (\mathrm{II}, 3)\} \tag{14}$$

Students who have already repeated level I advance to the next level with certainty. For ease of notation, I extend the definition to include the case of $\Pr(\text{fail}_{\mathrm{I},i2}) \equiv 0$. Moreover, I will sometimes refer to the "probability of graduation" in level II, defined as $\Pr(\text{grad}_{it}) = 1 - \Pr(\text{fail}_{\mathrm{II},it})$.

Vector w_{it} includes beliefs $\mathrm{E}_{i,t-1}(C_\tau)$, individual characteristics x_i, peer characteristics $p_{i,t}$ and the school leniency input $\mathcal{I}$. This specification accounts for the fact that there is no deterministic retention criterion, such that schools can choose to be more or less lenient. I assume that their degree of leniency in retention is proportional to their degree of leniency in grading, and therefore $\mathcal{I}$ is included as a regressor. Students are assumed to know the parameters and form expectations over their probability of graduation using (14).

While prior beliefs enter the retention probability, $C_{i,\tau}$ or equivalently h_i does not. This assumption appears to be reasonable since the school personnel do not know the student's true h_i either when deciding on retention, though they can form a belief about it by observing the student's performance, exactly as the student does. [27]

[26] For instance, using the previous notation, $I_{i,\mathrm{I}} = p'_{i\mathrm{I}} b_{p\mathrm{I}} + \alpha_{\mathrm{I}} \mathcal{P}_i$, for a student who did not repeat level I.

[27] Moreover, including $C_{i,\tau}$ in the model would be cumbersome, because students could then learn about their ability through the realization of the event. In fact, $\text{fail}_{\tau,it}$ would be a signal with binary value and a 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.

3.4 Educational choices

3.4.1 Flow utilities

Students receive a flow payoff for each period they spend in school. The payoff depends on beliefs about the level of their cognitive skills at the beginning of the period, on individual characteristics x_i, the history of retention ret_{it}, peer characteristics p_{it} and school inputs $\mathcal{J}$ and $\mathcal{M}_{\mathcal{M}}$ or $\mathcal{M}_{\mathcal{A}}$ (in the case of lower secondary education or upper secondary education respectively).

The flow payoff per period in lower secondary education for individual i attending level τ at time t is given by:

$$\begin{aligned} U_{it}^M &= \phi_{M,r}\mathrm{E}_{i,t}(C_\tau) + \text{ret}'_{M,it}\theta_{M,r} + x_i'\theta_{M,x} + p_{it}'\theta_{M,p} + \theta_{M,\mathcal{J}}\mathcal{J} + \mathcal{M}_M + \varepsilon_{it} = \\ &= \phi_{M,r}\mathrm{E}_{i,t}(C_\tau) + y_{it}'\theta_M + \varepsilon_{it}, \end{aligned} \tag{15}$$

The coefficients of the covariates capture all of the motivational and non-cognitive factors that affect the student's choice, in addition to their (perceived) level of cognitive skills. Moreover, the specification allows for differences among students according to retention history. In fact, the vector $\text{ret}'_{it} = (\mathit{stI2}_{it}, \mathit{stII3}_{it}, \mathit{ftII3}_{it})$ includes three mutually exclusive dummies to capture all of the possible combinations of time and retention. The baseline category is students who advance to level II at time $t = 1$. *stI2* takes a value of 1 if the student failed at $t = 1$ and has to repeat level I at $t = 2$. *stII3* takes a value of 1 if the student has to repeat level II at $t = 3$. *ftII3* takes a value of 1 for a student who repeated the first level at $t = 2$ and can advance to level II for the first time at $t = 3$. The value of the coefficient $\phi_{M,r}$ depends on the history of retention and as a result the level of skills may have differing effects on the flow utility, depending on the retention history.[28] For brevity, I will sometimes use $y_{it}'\theta_M$ for the total effect of the inputs other than the skill level.

I assume that $\mathcal{M}_M \perp \mathcal{J}$. Thus, it is convenient to define the total school input to utility as

$$\mathcal{T}_{\mathcal{M}} = \theta_{M,\mathcal{J}}\mathcal{J} + \mathcal{M}_M \tag{16}$$

In other words, $\mathcal{T}_M$ measures the overall school contribution to the educational decision beyond the school's effect through skills or retention. $\mathcal{M}_M$ is the component of the contribution that is unrelated to the school's grading policy.[29]

The flow utility for the choice of whether to enrol in high school has a similar formulation:

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

where $\text{ret}'_{A,it} = (rII_{it}, rI_{it})$, where $rII = 1$ if the student repeated the level II, and $rI = 1$ if they repeat level I. As above, the value of $\phi_{A,r}$ depends on the retention status. Finally, and as above, $\mathcal{T}_A = \theta_{A,\mathcal{J}}\mathcal{J} + \mathcal{M}_A$, where $\mathcal{M}_A \perp \mathcal{J}$.[30]

[28] 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.

[29] The assumption that $\mathcal{M}_M \perp \mathcal{J}$ is necessary for identification given that $\mathcal{J}$ is constant within a school.

[30] This formulation makes the implicit assumption that students take the grading policy $\mathcal{J}$ in middle school as a proxy for the grading policy in high school. It is worth recalling that students usually attend upper secondary

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.

3.4.2 Decision making

Students make an educational decision in each period, while taking in account their flow utility and expected future utility. Individuals are assumed to be forward-looking and choose the sequence of actions that yields the highest expected value of utility.

The one-period discount factor is δ. I use u_{it} to denote utility at time t. It is important to recall that students know all of the present and futures covariates described in the previous subsection, 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 level I or II. Therefore:

$$u_{it}|(\text{grad}_{it} = 1) = \max\Big\{0\,,\ U_{it}^{A}\Big\} = \max\Big\{0\,,\ v_{it}^{A} + \varepsilon_{it}\Big\}, \tag{18}$$

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

During lower secondary education. At $t = 2$, those who are still in school but have not yet graduated, again make the choice of whether to drop out, knowing that if they stay in school they will graduate with some probability and may gain access to upper secondary education.

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

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

education in the same school and under the same principal. Whether $\mathcal{I}$ has a significant effect on utility is an empirical question that will be answered by estimating the coefficient $\theta_{A,\mathcal{I}}$.

probability. Thus,

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

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

3.5 Identification

As is common for this type of dynamic discrete choice models (see, for example, 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 to 0.95 .[31]

The assumption that a student's ability affects educational choices only by way of its effect on cognitive skills is necessary in order to consistently estimate the parameters that govern skills and evaluations. Indeed, evaluations from time $t=2$ onward are observed only for individuals who chose to continue their education, raising a potential selection issue. However, econometricians can make the same inference as students do when they use the observed evaluations as signals for their own ability. Thus, they can compute the beliefs $\mathrm{E}(h)$ about ability and control for $\mathrm{E}(h)$ in order to consistently estimate the coefficients of the grade equations.[32] On the other hand, the noise of previous test scores, which is used as a signals to infer skills, only affects educational choices, but not skill development nor new test scores. This can be thought of as an exclusion restriction.

It is worthwhile to grasp the intuition of how the variation in the data is used to estimate the parameters of the grade equations. Assume for now that posterior beliefs $\mathrm{E}_i(h)$ have been computed for each student. Consider the following regression for external evaluations in level II at time $t\in\{2,3\}$, which mimics equation (12) in Subsection 3.2.5, except that time varying variables for previous levels are directly included as regressors:

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

where $z_{\mathrm{I},i}\widetilde{\kappa}_{\mathrm{II}} = \kappa_{\mathrm{II}} I_{\mathrm{I},i}$ and $z_{\mathrm{I},i}$ is the vector of time varying regressors for levels I and 0. Furthermore, $\widetilde{\epsilon}_{r_{\mathrm{II}},it} = \lambda_{\mathrm{II}}(h_i - \mathrm{E}_i(h)) + \epsilon_{r_{\mathrm{II}},it}$. Under the assumption that educational choices depend only on the student's belief about their ability, the error $\widetilde{\epsilon}_{r_{\mathrm{II}},it}$ is uncorrelated with the regressors (i.e. $(h_i - \mathrm{E}_i(h))$ is white noise) and therefore OLS can consistently estimate the parameters $o_{\mathrm{II}}, \beta_{\mathrm{II}}, \lambda_{\mathrm{II}}$.

Similarly, OLS can consistently estimate the reduced form parameters of the following regression

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

[32] Intuitively, $\mathrm{E}(h)$ is a good "proxy" for unobserved ability h because $h - \mathrm{E}(h)$ is purely white noise, and therefore uncorrelated with the other covariates.

(based on equation 13):

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

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

Having estimated γ, one can retrieve the structural parameters β_I and μ_I from an application of OLS 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, OLS applied to:

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

makes it possible 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 to retrieve the parameters κ_{I} and κ_{II}.

Up to this point, the identification of the parameters relied on the simplifying assumption that belief $\mathrm{E}_i(h)$ has already been computed. However, in order to perform the Bayesian updating one should know the variance σ of 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 parameters are identified from past signals and in particular the covariance between evaluations carried out in the same period or in different periods. In particular, σ can be inferred from the covariance between the residuals of $g_{\mathrm{II},it}$ and those of $r_{\mathrm{II},it}$, which are obtained by regressing the evaluation scores on the observable covariates. The variance of each type of residual is a linear function of σ and the variance of the relevant error, and therefore the latter can be retrieved after estimating σ.[33] In practice, the empirical application should jointly estimate the parameters that govern the grade equations and the covariance matrix for ability and signals.

4 Estimation

This section summarizes the main steps of the estimation procedure. Further details are provided in Appendix C.

Define $d_i = (d_{it})_t$ ($t \in \{1,2,3\}$) to be the vector of choices of student i, $\mathrm{fail}_i = (\mathrm{fail}_{it})_t$ to be the vector of retention/graduation events, and $o_i = (o_{it})_t$ to be the vector of evaluations observed by i.[34] The student makes $T_d \in \{1,2,3\}$ decisions, receives $T_{\mathrm{f}} \in \{1,2,3\}$ notifications of retention/graduation, and observes signals in $T_d + 1$ periods. More specifically, they receive T_d internal evaluations and $T_r \geq 1$ external evaluations. For instance, consider a student who is retained in level I, stays in school one more period and then drops out. They make two choices, i.e. $d_i = (1,0)$. At $t = 1$ they are retained, and at $t = 2$ they are promoted to level II, such that $\mathrm{fail}_i = (1,0)$.

[33] More precisely, $\mathrm{Cov}(g_{\mathrm{II},it}, r_{\mathrm{II},it} | z_{it}, z_{\mathrm{I},i}) = \mu\lambda^2_{\mathrm{II}}\sigma$. For instance, $\mathrm{Var}(r_{\mathrm{II},it}) = \lambda^2_{\mathrm{II}}\sigma + \rho^r_{\mathrm{II}}$.

[34] 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. I denote g_{it} (r_{it}) as the evaluation at time t, while abstracting from the level, and $g_{\tau,i}$ ($r_{\tau,i}$) as the last evaluation in level τ, while abstracting from time.

They are evaluated externally at the end of primary school, and internally in level I at $t = 1$ and $t = 2$, i.e. $o_i = (r_{0,i}, g_{\text{I},i1}, g_{\text{I},i2})$.

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

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

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

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

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 loglikelihood is separable into three parts (choices, retention probabilities, and evaluations) that can be estimated sequentially:

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

Estimation of $\log L_{i,o}$**.** Maximizing the likelihood $\log L_{i,o}$ would be computationally costly due to the integration of h. Following James (2011) and Arcidiacono et al. (2016), I use an Expectation-Maximization (EM) algorithm to overcome this issue. In a nutshell, it iterates two steps – the E-step and the M-step – until convergence. In the E-step of the kth iteration, the observed evaluation scores and the parameters estimated in the previous $k - 1$ iteration are used to compute the signals, estimate a posterior distribution $\psi_i^k(h)$ of each student's ability and update the estimate of the variance σ^k. The M-step maximizes the log-likelihood computed using individual posterior distributions and estimates a new vector of parameters.

Once the parameters have been estimated, I can compute beliefs about cognitive skills for each student at any point in time, and then use them as covariates in the subsequent stages.

Estimation of $\log L_{i,\textbf{fail}}$**.** I estimate the logit model for retention using the beliefs about cognitive skills and the previously found school leniency. Using the estimated parameters, I then compute the probability of repeating level I at time $t = 1$ and the probability of graduation in the subsequent time periods. These probabilities are then used in the next step.

Estimation of $\log L_{i,d}$**.** Given the beliefs and probabilities computed in the previous steps, I maximize the total log-likelihood in order to estimate the parameters, following Rust (1987). The main complication is that students use their beliefs about their skills in the flow utility, since they anticipate that if they stay in school they will receive new signals and therefore modify their beliefs. Thus, the computation of their expected utility for a given choice at time t requires integrating over all the signals that they may receive from $t+1$ onward.[35]

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, and for each school s, N_s individuals are sampled with replacements.[36] An alternative approach would be to not impose that the number of students in the school remain constant across iterations, and would simply draw with replacement N students from the total sample at each iteration. I estimated standard errors using this alternative approach as well and results were extremely similar (in particular the significance of the coefficients remained unchanged throughout).

5 Results

The main findings are summarized hereafter. Subsections 5.1 to 5.5 provide a more in-depth discussion of the estimated parameters.

Cognitive skills. About half of the total variance of cognitive skills is due to unobserved ability and about half to observed individual characteristics and school environment. Evaluations are informative: posterior individual variance diminishes rapidly and students soon acquire accurate beliefs about their skill levels.[37] Parental education is the most important observed determinant of cognitive skill accumulation. Having two parents with tertiary degrees as opposed to only basic education is associated with an improvement in skills by about 1 s.d. School environment is also quite important: higher quality peers increases performance, and there is large variation in school inputs. For instance, the difference between the school effect at the 75th percentile and the one at the 25th percentile (i.e. the interquantile range) is almost 0.4 s.d. School effects are homogeneous across individual characteristics.

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

Retention. School environment affects probability of retention and graduation beyond its impact on cognitive skills. For instance, increasing school leniency by 1 s.d. lowers the probability of retention by as much as an increase of 0.3 s.d. in cognitive skills. When choosing whether to pursue further education, retained students exhibit a flatter flow utility in cognitive skill level,

[35] This is the most computationally costly part of the maximum likelihood estimation and is performed using Gauss-Hermit quadrature.

[36] For two schools, in very few iterations the random sample did not includes any dropouts. In those instances, I resampled again the observations for the school. This happened in less than 4% of the iterations for one of the school and in less than 0.4% for the other.

[37] As explained in Section 3, the posterior individual variance is the updated variance of the student's belief after she receives one or more new signals.

and therefore there is a larger gap between retained and non-retained students at higher levels of cognitive skills than at lower ones. With regard to the choice of whether to continue on to high school, the value of the flow utility for retained students relative to their outside option is significantly lower than that for non-retained students at any given level of cognitive skills.

5.1 Cognitive skills and evaluations

Table 3: Variance of unobserved ability

	$\mu_l^2\widehat{\sigma}$	$\mathrm{Var}(C_l)$	%	$\mathrm{Var}(\mathrm{E}_l(h))$	$\mathrm{Var}(\mathrm{E}_l(C_l))$
$l=0$	0.280 (0.013)	0.572 (0.019)	49.0 (1.5)	0.112 (0.010)	0.406 (0.017)
$l=\mathrm{I}$	0.576 (0.018)	1.202 (0.068)	47.9 (2.4)	0.443 (0.018)	1.069 (0.084)
$l=\mathrm{II}$	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. $\mu_l^2\widehat{\sigma}/\mathrm{Var}(C_l)$. 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)
$t=1$	$r_0, g_{\mathrm{I},1}$	0.0650 (0.0040)
$t=2$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{I},2}$	0.0402 (0.0029)
$t=2$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}$	0.0392 (0.0028)
$t=2$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, r_{\mathrm{II},2}$	0.0321 (0.0019)
$t=3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{I},2}, g_{\mathrm{II},3}$	0.0286 (0.0022)
$t=3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{I},2}, g_{\mathrm{II},3}, r_{\mathrm{II},3}$	0.0246 (0.0016)
$t=3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, g_{\mathrm{II},3}$	0.0281 (0.0021)
$t=3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, r_{\mathrm{II},2}, g_{\mathrm{II},3}$	0.0243 (0.0016)
$t=3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, g_{\mathrm{II},3}, r_{\mathrm{II},3}$	0.0243 (0.0016)
$t=3$	$r_0, g_{\mathrm{I},1}, g_{\mathrm{II},2}, r_{\mathrm{II},2}, g_{\mathrm{II},3}, r_{\mathrm{II},3}$	0.0213 (0.0013)

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

The first two columns of Table 3 (Panel A) show, for each school level, the estimated variance of the individual's (unknown) ability h and the estimated total variance of cognitive skills. Given the assumption that h is uncorrelated with the regressors, it is possible to decompose it into variance due to observable characteristics, and variance due to ability. The third column of the table shows the share of total variance due to ability. This is the most interesting statistic because, as explained in Section 3, the variance of cognitive skills in a given level does not have any direct interpretation. In all levels, about half of the variance in cognitive skills is due to the variance in ability h. More specifically, the variance is about 49% before starting middle school and about 58% at the end of middle school.[38]

Panel B presents the posterior variance in each period for every possible set of signals. Evaluations are informative, and indeed the posterior variance is about 0.17 at time 0, it reduces to

[38] For the sake of comparison, the fourth and fifth columns of the table present the variance of beliefs about ability and about total cognitive skills in each level. In levels I and II, the variance of beliefs is similar to the previous estimates, although slightly smaller. This is consistent with the finding in the next panel, i.e. that students have accurate beliefs, and therefore the distribution of beliefs is similar to the distribution of actual ability.

0.06 at time $t = 1$, and to less than 0.04 at $t = 2$. It is about 0.02 for retained students who remain in middle school up to $t = 3$. Posterior variance is only slightly larger for students who did not receive an external evaluation in level II. In sum, when students make their first choice at time $t = 1$, they generally have relatively accurate beliefs about their ability and therefore about their cognitive skills. This suggests that their decisions would be unlikely to change if they were to observe their exact skill level rather than relying on their beliefs.[39]

Table 4 presents the estimated parameters governing evaluations, and in turn 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 presents 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 changes.

The results indicate that parental background is a fundamental determinant of cognitive skills. Having better educated parents is associated with large improvements in performance, with similar effect for internal and external evaluations. For instance, in level II having a mother with tertiary rather than primary education increases performance by more than 0.6 s.d., while having a highly educated father increases performance by more than 0.3 s.d. Thus, a student with two highly educated parents is expected to score almost 1 s.d. higher on internal and external evaluations than an identical classmate whose parents have only a basic education. Students whose parents completed high school (i.e. they have an "average" education level) are in an intermediate position: everything else equal they perform about 0.5 s.d. better than students with low-educated parents.

Immigrant students perform somewhat worse than natives. In contrast, there is no difference between cognitive skill accumulation of boys and girls. However, girls score higher on internal evaluations (by almost 0.4 s.d.). Being younger on entering primary education is a disadvantage, although the gap diminishes over time. Indeed, 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 at the beginning of middle school and with an increase of 0.14 s.d. at the end.[40] Starting middle school with a one-year delay is associated with lower performance (by up to -0.7 s.d. at the end of middle school). The quality of the student's neighborhood has a small positive effect on performance: an increase of 1 s.d.in neighborhood quality is associated with an improvement of at most 0.08 s.d.[41]

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

Repeating a level for a second time has a positive and sizable effect on cognitive skills. In particular, repeating level II is associated with a large improvement of about 0.5 s.d.

Peer quality has a positive effect on cognitive skills, while the proportion of females does not

[39] This intuition is confirmed in Section 6.4, in which student outcomes are compared to a counterfactual scenario without uncertainty.

[40] Calsamiglia and Loviglio (2019b) document the disadvantage of being younger at school entry in Catalonia, an effect that persists throughout compulsory education.

[41] The indexes of neighborhood quality and peer quality are described in Section 2.2. I also estimated the model with separate variables for a neighborhood's income level and education level, and the results are very similar. Therefore, I opted for the most parsimonious specification. Furthermore, I estimated the model with a vector of variables for peer characteristics, results are aligned although less precisely estimated.

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 $\mathcal{I}$): 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.

appear to have any relevant effect. In order to capture any non-linear peer effect, I use a polynomial of degree 3 for each peer regressor. To facilitate the interpretation of the results, Table 4 shows the effect associated with having peers at the mean versus 1 s.d. below the mean, which is associated with an improvement of almost 0.3 s.d. in level I. Figure A-11 in Appendix A.2 plots peer effects along the distribution of the peer variables. As can be seen, better peers always increases evaluations in level I, although the effect is more pronounced in the tails of the distribution. The pattern is similar in level II, except that the effect is smaller in magnitude, though still significant. On the other hand, the positive effect is offset in internal evaluations. This is in line with the finding in Calsamiglia and Loviglio (2019a) that in Catalonia teachers' evaluations are deflated in the presence of better-quality peers, i.e. teachers are "grading on a curve" to some extent. The proportion of female peers has no effect on skills at any point in the distribution.

With respect to the estimated school inputs $\mathcal{A}$, Table 4 presents the difference between the 75th percentile and the 25th percentile (interquantile range), between the 80th and 20th percentile, and the standard deviation. $\mathcal{A}$ varies considerably across schools. For instance, the interquantile range is almost 0.4 s.d. in both level I and II, which is comparable to previous results in the literature. For example, replacing the lowest-ranked Boston public school with an average school would improve performance of the affected students by 0.37 s.d. (Angrist et al., 2017). Similarly, attending an oversubscribed charter school in Boston improves performance by approximately 0.4 s.d. in math and 0.2 s.d. in English (Abdulkadiroğlu et al., 2011).

Table 4 presents similar statistics for school input $\mathcal{I}$, which captures school grading policies. Its variation across schools is also sizable. Finally, school inputs $\mathcal{I}$ and $\mathcal{A}$ are negatively correlated (correlation of about -0.7). In other words, schools that improve cognitive skills the most tend to have a stricter grading policy for internal evaluations.[42]

The empirical specification includes interactions between the vectors of school inputs $\mathcal{A}$ and $\mathcal{I}$, and the dummies for gender, immigrant status, and parental education, which is intended to account for variation in school effects across socioeconomic groups. If, for instance, the school's effecton cognitive skills is larger for girls than for boys, then this would be captured by the coefficient of the interaction term. To ease the interpretation of the results, Table 4 reports the change due to an increase of 1 s.d. in the school effect for each individual characteristic. The estimated interaction effects are small and insignificant. Therefore, the effect of school environment on cognitive skills development is about the same for all students enrolled in a given class, regardless of gender, nationality or parental education.

5.2 Retention and graduation

Table 5 presents the estimated parameters governing the logit model for retention.[43] Unsurprisingly, higher cognitive skills decreases the probability of retention. Girls have a lower probability of retention regardless of skill level. Parental education appears to be less relevant, although having highly educated parents (especially a father) is associated with a negative effect.[44]

[42]This is consistent with the the evidence presented in Calsamiglia and Loviglio (2019a) that schools in Catalonia with higher average external evaluation scores have stricter grading policies.

[43]Table 5 has the same structure of Table 4.

[44]A student's behavior in class may vary with parental background, and it may affect the retention decisionn. Moreover, teachers communicate with parents during the school year if their child is at risk of retention, and to some

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.

As shown by Figure A-11, having higher quality peers increases the probability of retention at any given level of skills. This is consistent with the negative effect of peer quality on internal evaluations discussed in the previous subsection. Teachers may be more prone to fail a student who lags behind his peers than an identical student who is close to the median of his class. Furthermore, students in more lenient schools (larger $\widehat{\mathcal{I}}$) are less likely to be retainedand the effect is sizable. For instance, increasing school leniency $\mathcal{I}$ by 1 s.d. has about the same effect on retention as increasing the school input $\mathcal{A}$ by 1 s.d. (thereby improving skills by almost 0.3 s.d.).[45]

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 not to drop out and the choice to enroll in high school.

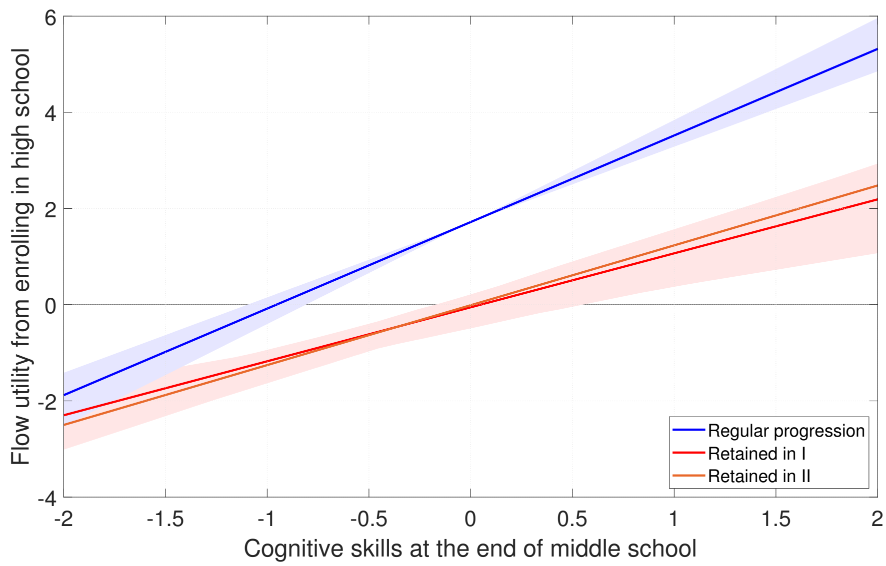
extent they take into account the parents' preferences. This may explain the results.

[45]Results are similar when the leniency of the school's retention policy is estimated directly using school dummies. The correlation between the estimated vector of school effects and $\mathcal{I}$ is more than 0.8. Given that rules for grading and for retention are most likely designed jointly at the school level, I prefer to estimate only one parameter for each school.

Girls are less likely to drop out and more likely to enroll in high school after graduation. Students with a highly educated father are more likely to choose to enroll in high school.

Retention during lower secondary education has a negative effect on flow utility from the choice of high school. The interactions of cognitive skills with dummies for repeating level I or level II also have negative coefficients, and therefore the total coefficient of cognitive skills for retained students is smaller than for students who are progressing regularly.[46] In other words, retention decreases the weight that individuals attribute to cognitive skills when making their choices. Figure 5 helps to illustrate this point by showing that, at any level of skills, retained students have a lower payoff, although the gap is wider at higher levels of cognitive skills. Thus, an increase in skill levels has less of a positive effect on the utility for retained students than for non-retained students.[47] As discussed in Subsection 5.2, retention likelihood is decreasing in cognitive skill level, but it is still high for average or just below average skill level, especially among male students. Retention may in particular discourage these students from pursuing higher education. The effect of retention on the choice to stay in school goes in a similar direction, although the coefficients are generally smaller in size and not precisely estimated.

Figure 5: 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".

For a given level of cognitive skills, students with below-average peers are more likely to choose

[46]The coefficient of the interaction with the dummy for repeating level I is significant at 1%, while the other coefficient 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 regular and retained students would be captured by the coefficient of the dummies. Thus, the fact that the lines for retained students lie below the one for regular students does not mean that they like high school less in absolute terms, but rather that they value it less compared to their outside options.

to stay in middle school than those with above average peers. This may be related to ranking concerns: students at the margin of dropping out may have a further reason to leave school if they dislike being at the bottom of the class.[48] Figure A-11 shows that the effect is non-linear: it flattens for values of peer quality above the average. No relevant peer effect is detected for the choice to enroll in high school.

Total school inputs $\mathcal{T}_M$, which affect the choice to stay in middle school, and total school inputs $\mathcal{T}_A$, which affect the choice to enroll in high school are sizable. In both cases, school input at the 75th percentile rather than the 25th percentile increases the flow utility by as much as improving cognitive skills by 0.5 s.d. Both $\mathcal{I}$ and the orthogonal inputs $\mathcal{M}_M$ and $\mathcal{M}_A$ explain a large portion of the aggregate school inputs on choices. About 50% of the variance of $\mathcal{T}_M$ is due to $\mathcal{M}_M$, which increases to 75% for $\mathcal{T}_A$ and $\mathcal{M}_A$. The direct effect of school inputs on the flow utility from high school is comparable to the indirect effect through cognitive skills. Indeed, raising either $\mathcal{M}_A$ or $\mathcal{A}$ by 1 s.d. increases the flow utility by about 0.5 s.d.[49]

It is important to stress that although school inputs have a linear effect on flow utility, the school environment may differentially affect classmates according to their individual characteristics or cognitive skills. This is due to the non-linearity of the probability functions. such that the effect on the probability of a change in one of the regressors depends on the initial level of the underlying utility.[50] Thus, the school environment may differentially affect the educational decisions of students with differing background even if they are in the same class. This will be discussed extensively in the following sections.

5.4 Fit of the model

To assess the fit of the model, I simulate the choices and outcomes of each individual in the sample, using the structural parameter estimates presented in the previous sections. More specifically, I create 1000 copies of each individual at time 0 (according to their time invariant characteristics, the primary school they attended, and the middle school they are attending). For each of them, I draw ability and shocks to evaluation scores, preferences, and retention events, using the estimated distributions. I can then compute their outcomes, and in particular their cognitive skills and choices.

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

[48]The findings in Pop-Eleches and Urquiola (2013) corroborate this interpretation. They find that having better peers has a positive effect on achievement, but children who make it into more selective schools realize they are relatively weak and feel marginalized.

[49]The effect of increasing $\mathcal{A}$ is the product of the increase in skills (as in Table 4) and the effect of skills on flow utility (as in Table 6).

[50]As explained in Section C.3, given the assumption that shocks to preferences follow a logistic distribution, the probability of making 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 expected utility (i.e. the flow utility at time t before observing the shock to preferences plus future expected). The marginal effect of a change in the regressor z_k is

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

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

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.

the model (in the "model" columns) for the following events: the choice to stay in school at time $t = 1$, graduation, enrollment in high school, retention in level I, retention in level II. Frequencies are computed over the entire sample at time $t = 1$. Table A-11 in the appendix reports similar statistics calculated only for the subsample of the initial population who reached the relevant stage for the event to take place (e.g. enrollment in high school is calculated for the subsample of students who completed middle school). The first line of each table presents frequencies for the overall sample, while the subsequent rows present that same information by subgroups (such as parental education, gender, etc). The predicted choice of whether to stay in school, graduation rate, and choice of whether to enroll in high school are very close to the actual ones, both in the overall sample and by subgroup. The retention rate in level I is also almost identical to the actual one, while the retention rate in level II is slightly higher.

I next investigate how evaluations and events simulated by the model replicate the patterns observed in the data. For instance, Figure A-12 plots the proportion of students who chose to stay in school at $t = 1$ by test score quantile at $t = 0$. Figure A-13 plots their enrollment in high school, again by test score quantile. The model predictions successfully mimic the empirical outcomes. The other evaluations and choices exhibit similar patterns.

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.

5.5 Summary of school inputs by parental background

The results presented in this section show that the school environment, namely peer quality and school inputs, has a sizable effect on cognitive skills, on the logit function that determines promotion and retention, and on the flow utility of educational choices. The overall effect of the school environment on attainment depends on the interplay between all of these channels. To understand this interplay, in the next section I will simulate the outcomes of several groups of students when exposed to different school environments. In particular, I will focus on differences by parental education. At this point, it is worth summarizing the results regarding the effect of the school environment by parental education. According to the results in Section 2.3, students with highly educated parents have on average higher quality peers than students with low-educated parents.

However, it is important to recall that all the schools in the sample have a diverse pool of students with respect to parental education. Moreover, on average, students with low and high parental background are exposed to similar school inputs. As shown in Table 8, the average grading policy $\mathcal{I}$ is almost identical for the two groups. $\mathcal{A}$ and $\mathcal{M}_M$ are slightly larger for students with low-educated parents, but the differences are small in magnitude.[51] The only exception is the effect of school input on the choice of whether to enroll in upper secondary education: students with low-educated parents are more likely to attend schools that encourage high school enrollment (such that $\mathcal{M}_A$ is significantly larger in their case). Students with high and low parental education can be found

Table 8: Average school input by parental education

	$\mathcal{A}$	$\mathcal{I}$	$\mathcal{M}_M$	$\mathcal{M}_A$
low parental edu.	0.019	0.014	-0.007	0.280
high parental edu.	-0.032	0.021	-0.069	-0.228
p-value difference	0.15	0.847	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.

in all schools in Barcelona, and those schools vary widely in the inputs they provide. Therefore, it is important to study how the effect of school environment on student outcomes varies with a student's background.

6 Variation in educational outcomes across schools

Using the estimated parameters of the model, I simulate the educational outcomes that every student in the sample would have in each middle school in Barcelona. More specifically, I simulate student i's outcomes in school s using i's individual characteristics and test score at time $t = 0$ and their estimated unobserved ability, Peer characteristics are set at the average value of each school and level. For each student i and school s, I draw 1000 sets of shocks to evaluations, preferences, and retention events using the estimated distributions, and simulate i's choices and outcomes. Averaging the results, I then compute i's expected outcomes in school s. More specifically, I focus on three main outcomes of interest: 1. cognitive skills C_{II} at the end of lower secondary education; 2. the ex ante (at time $t = 0$) probability of graduation; and 3. the ex ante probability of enrolling in academic upper secondary education.[52] The first outcome is a measure of performance, while the other two are measures of attainment (for brevity, I will refer to them as "graduation" and "high school enrollment", respectively). Analogous aggregate measures are computed by averaging – at the school level – the observed outcomes in the data: 1. average external evaluations at the end of lower secondary education; 2. the proportion of graduates; and 3. the proportion of students who enroll in high school.

[51] They are only significant at the 10% for $\mathcal{M}_M$.

[52] Skills at the end of lower secondary education are computed using the occurrences in which the student reacheslevel II. This occurs in at least some iterations for all students in all schools.

In Subsection 6.1, I compute three school rankings for each student based on the three simulated outcomes of interest and compare the individual rankings to the aggregate rankings from the observational data. Furthermore, I analyze the correlation between school rankings based on different outcomes. In particular, I assess whether school rankings based on performance are a good proxy for school rankings based on attainment. In Subsection 6.2, I quantify differences across schools for specific subsamples of the population, in order to determine who would particularly benefit from attending a highly ranked school according to performance or attainment. Subsection 6.3 extends the analysis by highlighting the role of innate ability and parental education. Subsection 6.4 focuses on the effect of retention on attainment. Subsection 6.5 examines a counterfactual policy that makes education compulsory until the end of middle school is reached.

6.1 Student-specific school ranking

Table 9: School Rankings

	C_{II}	graduate	enroll in high school
avg rank corr(data, simulation$_i$)	0.891 (3.7e-04)	0.554 (0.002)	0.594 (0.001)
avg sd(rank$_s$)	1.059 (0.099)	6.037 (0.289)	4.462 (0.335)

	Corr(C_{II}, grad.)	Corr(C_{II}, high sch.)	Corr(grad., high sch.)
rank corr in the data	0.646	0.778	0.897
avg rank corr in the simulation	0.151 (0.002)	0.233 (0.002)	0.691 (0.002)

Note. The first line of the top panel plots the average correlation between individual rankings simulated through the model and observed data for the three educational outcomes displayed in the columns (skills at the end of lower secondary education, graduation, enrollment in high school). The second line plots the average standard deviation of the school position across individuals.
The second panel of the tables plot correlations between rankings based on different outcomes; the first line shows the empirical correlations, the second line the average correlations between individual rankings. Standard error of the mean in parentheses.

I first look at rankings separately by outcomes. For each student, I compute the correlation between each individual ranking and its aggregate counterpart from the observational data. Mean correlations are shown in the first line of Table 9. Moreover, I assess how school rankings for a given outcome vary across students by calculating for each school the standard deviation of the school ranking across students.[53] Average results are shown in the second line of the table. On average, the correlation between individual and aggregate rankings by performance is high (0.89). Moreover, individual rankings have a low variance: on average, the standard deviation of the position of a given school across individual rankings is about 1, which corresponds to about 2 percentiles on a 0-100 scale. Conversely, the table shows that the correlation between individual and aggregate rankings is positive but lower for attainment than for performance (0.55 for graduation and 0.59 for high school enrollment, respectively) and there is sizable variation across individual rankings.

[53] If a school is ranked the same for all students in the sample, the standard deviation is 0. Conversely, if the school's position follows a discrete uniform distribution (i.e. it can take any position with equal probability) its standard deviation is $\sqrt{(\#\text{schools}^2 - 1)/12}$, which is equal to almost 14 when there are 47 schools.

On average, the standard deviation of the school ranking is 6 (13 percentiles) for graduation and 4.5 (10 percentiles) for high school enrollment.

Of interest is also the cross-correlation between school rankings according to different outcomes. As discussed in Section 2.3, average performance and average attainment at the school level are positively correlated in the data. More specifically the rank correlation between observed performance and graduation rate is 0.65 and while that between performance and high school enrollment rate is 0.82. This relatively large correlation in the data may be driven by the characteristics of the students. For instance, schools with a high proportion of students with a high socio-economic background typically show both high test scores and high attainment. Indeed, results in the second panel of Table 9 show that the correlation between performance and attainment plummets when individual rankings are used. It drops to 0.15 for graduation and to 0.23 for high school enrollment.

The results up to this point suggest that a simple school ranking based on average test scores can provide useful information about a school's contribution to skill development, and therefore performance – provided that the student does not drop out. On the other hand, it does not convey much information about the student's attainment prospects. This is because the correlation observed in the raw data at the school level between performance and attainment is driven mainly by student characteristics, rather than a school's impact on attainment. In particular, it is not informative about how a given school's environment would contribute to the attainment of students who are not typical of their current student population.

6.2 Differences in outcomes by student characteristics

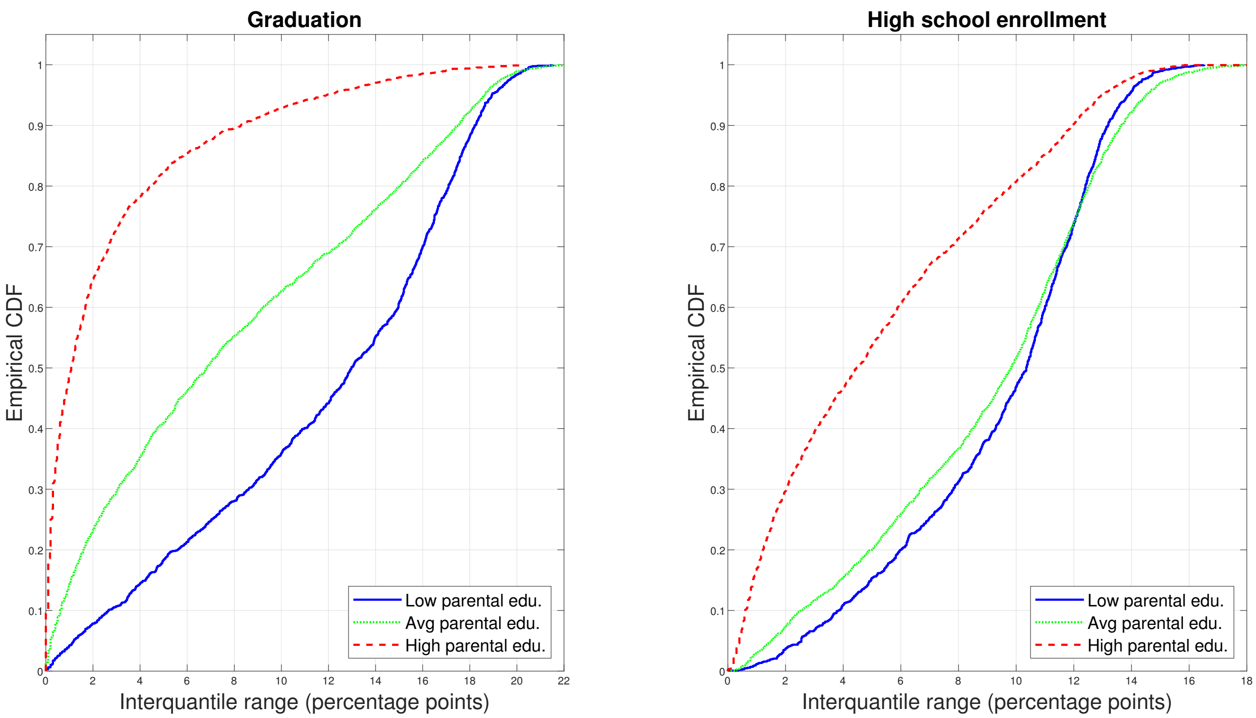
Ignoring attainment is particularly problematic in the case of students whose probability of graduation or high school enrollment vary significantly across schools. While the discussion so far has focused on ordinal rankings, in this subsection I assess the magnitude of a school's contribution to the educational outcomes of students according to parental education.

More specifically, for each student I measure the variation of simulated outcomes across schools using the interquantile range (IQR), namely the difference between attending the school at the 75th percentile of the outcome distribution and attending the school at the 25th percentile. I find the interpretation of the IQR to be particularly intuitive, nonetheless I also repeat the analysis using other measures such as standard deviation.

Figure 6 plots the empirical cumulative distribution of the IQR of graduation probability (left panel) and the IQR of high school enrollment probability (right panel) for different levels of parental education. Table A-12 in the appendix presents the average IQR and standard deviation in the entire sample and by parental education.

The results confirm the intuitions discussed in Section 5 and make it possible to quantify them. Skill levels at the end of lower secondary education vary considerably across schools (average IQR of about 0.40 s.d.). Given that the production function of skills is almost linear in its inputs, the contribution of the school environment to skill development is about the same for every student enrolled in a given school with a given set of peers. The variation across schools is also sizable for graduation and high school enrollment: the average IQRs are about 7 p.p. and 8 p.p., respectively. However, it is evident from Figure 6 that there are major differences across groups. In fact, students with lower parental education improve their skills less over the years, therefore they are more likely

Figure 6: Variation of outcomes across schools. CDF by parental education



Note. The figure plots the empirical CDF of the interquantile range of the graduation probability (left) and high school enrollment probability (right) by parental education (low, average, and high). The interquantile range is the difference between the outcome in the school at the 75th percentile of the distribution and the school at the 25th percentile. Statistics are computed using student simulated outcomes in each school.

to repeat a level and drop out. Therefore, on average they have a lower probability of graduating and enrolling in high school than students with higher parental education (because of both their skill level and the direct negative effect of retention on educational choices). Given that they are at the margin between staying and leaving, the school environment has a larger effect on their probability of pursuing further education.

The empirical CDF in the graph shows that the median IQR for graduation is about 13 p.p. for students with low-educated parents, while it declines to about 1 p.p. for students with highly educated parents. Moreover, the IQR is more than 9 p.p. for 70% of the students with low-educated parents and 16 p.p. or more for 30% of them. Conversely, it is less than 3 p.p. for 70% of the students with highly educated parents, and less than 0.5 p.p. for 30% of them. The empirical CDF for students with average parental background lies approximately midway between those two groups. The situation is comparable for high school enrollment, though with two main differences. First, the CDF for average parental background is closer to the one for low parental background. Second, the variation across schools is non-negligible for many students with high parental education as well, such that the median IQR value is almost 5 p.p.

Table A-13 in the appendix repeats the analysis by gender and immigrant status. The variance of attainment outcomes across schools is somewhat larger for males than females (for example, the average IQR for graduation are 8.2 p.p. for males and 6 p.p. for females). This is consistent with the results in the previous section which showed that, everything else equal, males are more at risk of failing a level and dropping out. Therefore, the contribution of school environment is more likely to make a difference for their choices. Finally, attainment varies more across schools for immigrant students than for natives.

6.3 The interplay between school, ability and parental background

Students in the sample who differ according to parental education may also differ along other dimensions. For instance, students with low-educated parents are more likely to be immigrants and to live in a poorer neighborhood. Therefore, I repeated the analysis described in the previous subsection for two representative types of students, who are identical in every observable and unobservable characteristics except parental education. *Type L* has low-educated parents, and more specifically both parents attained at most lower secondary education. *Type H* has highly educated parents, and more specifically both parents completed tertiary education.[54] The following are individual characteristics shared by both types: male, native Spanish, born in the middle of the year, began lower secondary education at 12 years of age, and attends the first-choice middle school. Primary school effect, cohort effect, and neighborhood quality are all set to their average value.[55]

I simulate their outcomes in every school for values of innate ability ranging from about -3 s.d.

[54] Overall, 18% of students in the sample have two parents with at most lower secondary education. In the case of the others in the group with low-educated parents, one parent has basic education and the other education level is unknown. 16% of the students have two parents with tertiary education. In the case of the others with highly educated parents, one parent has tertiary education and the other has upper secondary education or it is unknown.

[55] I repeated the analyses with other sets of characteristics, particularly for female and immigrant students. Overall, the results exhibit similar patterns, and the differences go in the expected direction. For instance, females are less likely to be retained, and they more often choose to pursue further education, while immigrants are more likely to drop out.

below average to about +3 s.d. above average.[56] It is important to stress that in all the simulations type L and type H are exposed to exactly the same environment in every school. This makes it possible to quantify the relevance of school environment for each type at a given level of ability. Figure 7 plots the attainment of type L (blue lines) and type H (red lines) in the school at the 75th percentile (continuous lines) and in the school at the 25th percentile (dotted lines). Figure A-14 in the appendix shows the interquantile ranges.

The left panel shows that school environment has the greatest impact on graduation probability for type L students with innate ability about 1 s.d. below average. For them, attending the school at the 75th percentile rather than the one at the 25th percentile improves their graduation rate by up to 18 percentage points, i.e. by more than 40%. The IQR is still large (about 12 p.p.) for average ability and gradually decreases towards 0 as ability increases. Conversely, the school environment has a sizable effect on type H's probability of graduation only for very low-ability students.

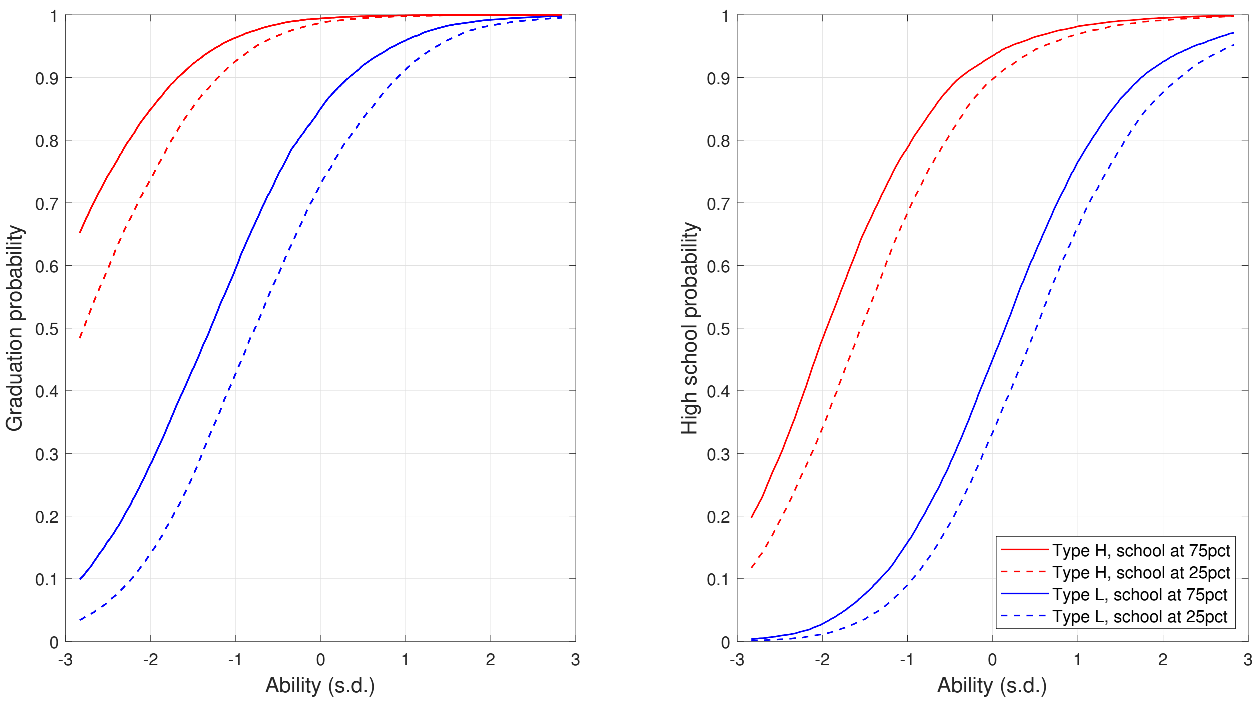
The right panel shows that the probability of enrolling in high school is particularly sensitive to the school environment for type H students with below-average ability and for type L students with average or above-average ability. For both groups, being in the school at the 75th percentile rather than in the one at the 25th percentile increases the probability of enrolling in high school by up to 14 p.p., an improvement of more than 30%.

Table A-14 in the appendix focuses on two students with average ability ($h = 0$) and shows the variation across schools of skill levels in level I and level II, retention, drop out, graduation and high school enrollment (conditional on graduation and otherwise). Herein, the type L and type H students with $h = 0$ will be referred to as L_0 and H_0, respectively. Figure 8 shows the main outcomes of interest for each school in the sample. In the left panel, the y-axis measures the probability of graduation, and the x-axis measures the cognitive skills that L_0 (blue dots) and H_0 (red diamonds) would achieve in each school if they successfully graduate. The right panel shows the same for the probability of enrolling in high school. While the variation of cognitive skills across schools is similar for both types, the probability of graduation for H_0 is above 94% for all schools and close to 1 for all schools with above average expected cognitive skills, and the probability of enrolling in high school is above 80% for all but one school. Meanwhile, L_0 exhibits greater dispersion across schools: the probability of graduation ranges from 55% to 90%, and the probability of enrolling in high school ranges from less than 20% to about 60%.

Figure 8 visually summarizes the main findings discussed in this section. First, the correlation between school contribution to performance and attainment is positive but not high: many schools that achieve an average level of skills ensure higher attainment than schools that have a much larger effect on skills. Second, this a is crucial factor for less advantaged students. Thus, for the average student with high parental education, the x-axis alone conveys the most salient information for assessing the impact of attending a given school. On the other hand, multiple dimensions are necessary to understand the contribution of schools to the educational outcomes of the average student with low parental education.

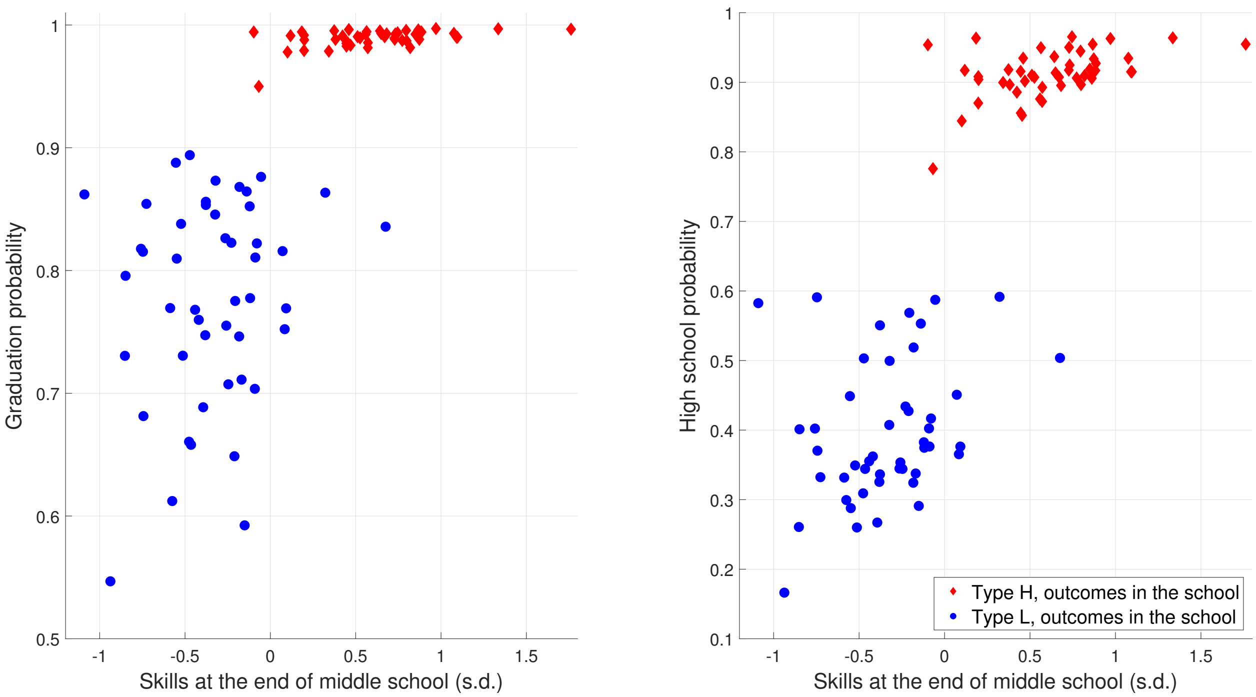
[56] As discussed in Section 5.1, the estimated variance of ability is about 0.28. Therefore, I use values ranging from -1.5 to 1.5 in the simulation. The graphs show standardized values in order to facilitate the interpretation of the results.

Figure 7: Attainment of type L and type H students by ability



Note. The left figure plots the graduation probability (y-axis) by ability level (x-axis) for four representative student types: student with highly educated parents enrolled in a school with high graduation probability (continuous red line), student with highly educated parents enrolled in a school with low graduation probability (dotted red line), student with low-educated parents enrolled in a school with high graduation probability (continuous blue line), student with low-educated parents enrolled in a school with low graduation probability (dotted blue line). The school with high (low) graduation probability is the one at the 75th percentile (25th percentile) of the school ranking by graduation probability. Each probability is the average of 1000 simulations for a given student type at a given level of ability. For all simulations the individual characteristics are: male, Spanish, born on July 1, began middle school at 12 years old, attends the middle school at the top of the application list; primary school effect, cohort effect, neighborhood quality are set at their average values. The right figure has the same structure, but the outcome used is the probability of enrolling in high school.

Figure 8: Outcomes of type L_0 and type H_0 by school



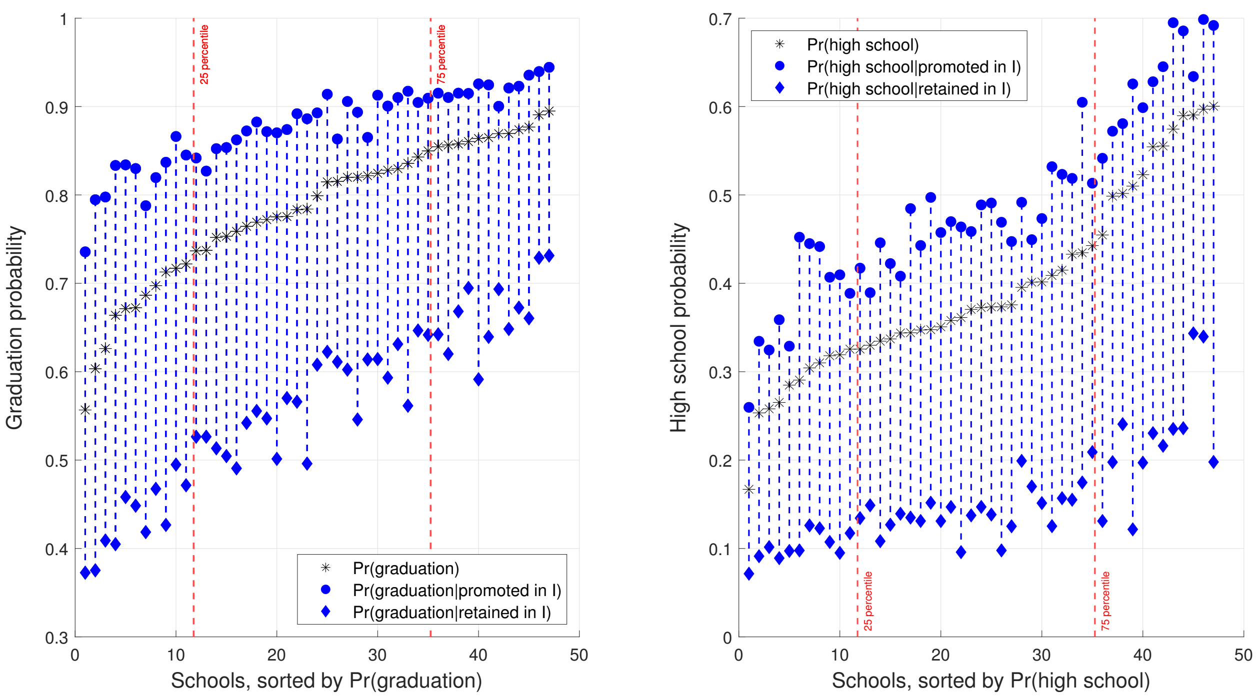
Note. The left figure plots graduation probability (y-axis) and skills at the end of lower secondary education (x-axis) in each school of the sample for two representative students: Type L_0 (blue dots) has parents with primary education, type H_0 (red diamonds) has parents with tertiary education. hey both have average ability $h = 0$, and are male, Spanish, born on July 1, they began middle school at 12 years old. Primary school effect, cohort effect, and neighborhood quality are set at their average value. The average peer characteristics in each school are used. Frequencies are constructed using 10000 simulations of the structural model for each type and every school in the sample. The right figure has exactly the same structure, but the outcome on the y-axis is the probability of enrolling in high school.

6.4 Retention and its consequences

In Barcelona, one of every three students with low parental education is retained during lower secondary education. Correspondingly, the L_0 student introduced in the previous subsection has about a 30% average probability of being retained in level I. Depending on the school in which he is enrolled, the probability ranges from slightly less than 20% to almost 50%.

The results presented in Section 5 show that repeating a level eventually improves a student's skills, while the overall effect of retention on attainment is ambiguous. On the one hand, higher skill levels improve the flow utility from schooling and make failure less likely in the future. On the other hand, retention may affect attainment negatively. There are three main reasons for this. First, every thing else equal, retention occurs more frequently in less lenient schools: stricter grading policies have a negative direct effect on flow utility and decrease expected utility because they make failure more likely in the future. Second, repeating a level has a sizable direct negative effect on flow utility. Third, the noise of the evaluations affect both the retention events and beliefs formation: retained students may underestimate their skills, which in turns will negatively affect their choices.

Figure 9: Attainment of student L_0 in each school by retention status.



Note. The left figure plots graduation probabilities by school for a representative student with average ability and low-educated parents. The figure shows ex-ante probabilities at $t = 0$ (black stars) and probabilities conditional on promotion at $t = 1$ (blues dots) or retention at $t = 1$ (blue diamonds). The vertical red dotted lines shows the 25 and 75th percentiles in the distribution of graduation probability. The right figure follows the same structure, using enrollment in high school as outcomes. These graphs are produced using the simulation described in Figure 8.

As in the previous subsection, I focus on L_0 and study how retention affects his outcomes. Figure 9 compares L_0's attainment in two scenarios: he is promoted at time $t = 1$ (blue dots) versus he is retained at time $t = 1$ (blue diamonds). Schools are sorted according to overall probabilities along the x-axis (plotted as black stars).[57] It is important to recall that only the random errors vary in the simulation, while $h = 0$ and all the observable characteristics are constant across iterations. In particular, given the school environment, L_0 always has the same skill level at $t = 1$, and therefore whether he is retained depends on the noise of the evaluations (by way of the beliefs about skills) and on the shock to the retention logistic function.

Differences between the two scenarios are striking for all schools: both graduation probability and high school enrollment drop by between 20 p.p. to 50 p.p. when L_0 is retained. Retention amplifies differences in graduation probability across schools: L_0 rarely drops out after successfully advancing to the next level, and thus in almost all schools he has more than an 80% probability of graduating. Conversely, retention reduces the differences across schools in high school enrollment: L_0's chances of reaching upper secondary education are between 10% and 30% if he fails level I, while they range from about 30% to 70% if he is promoted.

This graphs confirms that the overall effect of retention on attainment is negative and sizable for all schools, regardless of grading policy. It is worthwhile at this point to explore other channels through which retention can affect attainment, while focusing on the outcomes of L_0 when enrolled in a "representative" school. In the simulation, peer variables and school inputs are set at the average values among students with low-educated parents. Table A-16 in the appendix presents the results of the simulation. In particular, it compares beliefs and true cognitive skills at each point in time by retention status. At time $t = 1$ and $t = 2$, the student's beliefs are lower than his actual skill level if he is retained, although the difference is small. Thus, he would be only slightly more likely to drop out even in the absence of any direct negative effect of retention on utility. Conversely, if he stays in school and graduates at time $t = 3$, his perceived (and actual) skill level is higher than his perceived skill level after graduation at time $t = 2$. Thus, the large gap in high school enrollment is due to differences in preferences, rather than beliefs.[58]

To quantify the impact of uncertainty about one's ability on choices and outcomes, I repeat the analysis under a counterfactual scenario in which student ability h is perfectly known instead than unobserved. The results in Table A-17 in the appendix show that removing uncertainty about ability would increase L_0's graduation probability by about 2 p.p. and his enrollment in high school by about 0.5 p.p. This is primarily due to a lower likelihood of dropping out after retention, both at time $t = 1$ and at $t = 2$. Conversely, the rate of enrollment in high school would overall remain almost unchanged. Specifically, it decreases slightly if the student graduates at $t = 2$ (given that he somewhat overestimated his ability) and it increases slightly if the student graduates at $t = 3$. Overall, retained students are somewhat penalized by the randomness of the signals, although most of the difference in choices and attainment is due to preferences.

Finally, it is noteworthy that in both scenarios L_0 is very likely to graduate if he does not choose to drop out. While his probability of not completing lower secondary education is as high as 23%, only 4 p.p. are due to failure at $t = 3$. The remaining 19 p.p. are due to dropping out at time

[57] Computations are based on the simulation described in Subsection 6.3 (Figure 8 and Table A-14). Table A-15 in the appendix reports IQR and s.d. across schools for graduation, enrollment in high school, dropping out at $t = 1$, and enrollment in high school conditional on graduation.

[58] I replicate the analysis for student H_0. His outcomes follows a similar pattern. In fact, although he still has much better prospects than type L_0 even after being retained, retention is associated with large drops in graduation and enrollment rate. However, it is worth to recall that retention is very unlikely for this student (about 3% in first level and only 1% in second level).

$t = 1$ or $t = 2$.

6.5 A counterfactual policy for mandatory education

In Spain lower secondary education is often thought of as compulsory, given that students normally turn 16 years old during their last year in middle school. However the previous analyses made it clear that in practice the legal dropout age is not necessarily aligned with the end of lower secondary education. I therefore use the model to explore the effects of a counterfactual policy, in which school attendance is mandatory until a given level, rather than a given age. More specifically, students are required to stay until the end of level II, repeating a grade if necessary; they are not allowed to leave at $t = 1$ or $t = 2$, unless they graduate. Those who fail at $t = 3$ fulfill the legal requirement, but, as in the baseline model, cannot enroll in upper secondary education.

It is important to acknowledge that such a policy change might prompt schools to respond by somehow adjusting their inputs. However, this goes beyond the scope of the exercise, which treats school inputs as given. Nonetheless, the results illustrate what would happen in the short run and can be taken as a benchmark for the outcomes absent any school adjustment.[59]

As in the previous simulations, I create 1000 copies of each individual in the sample and, for each copy, I draw ability and shocks to evaluations, preferences, and retention events, using the estimated distributions. I can then compute student outcomes under the baseline and the counterfactual scenarios. Panel a) of Table 10 summarizes average outcomes in the population and by subgroups of parental education. For comparison, panel b) replicates the exercise for the various student types: L_0, M_0, H_0. As before, individual characteristics other than parental education are kept constant and ability is set at its average value of 0. Each type is assigned the average school environment for students with similar parental education.

Changing the rules for mandatory education would increase the overall graduation rate to 94%. This is consistent with the finding in previous subsections that the overwhelming majority of students would graduate if they stayed in schools until $t = 3$. The increase is particularly large for students with low-educated parents (+17 p.p.), but it is also sizable for students with average parental background (+10 p.p.). Thanks to the large increase in the pool of middle school graduates, overall enrollment in high school increases by almost 4 p.p. (6 p.p. in the case of students with low-educated parents), while enrollment conditional on graduation slightly decreases, since students at the margin of dropping out are not very likely to pursue further academic education if they eventually graduate. The results for selected student types are consistent with these patterns in the general population.

Making education compulsory up to a specific level rather than a specific age does not have any major effect on the average skill level of high school students, which decreases by only 0.07 s.d., and therefore it appears unlikely that such a change would have any notable effect on the next educational stage by way of a change in peer composition. In the case of middle school graduates, the decrease is modest though slightly larger (about 0.13 s.d.). Overall, the results provide further

[59] On the other hand, I allow peer effects to vary in the counterfactual simulation. More specifically, peer quality usually worsens. In fact, in the baseline case, peers in level II are somewhat positively selected, given that many students drop out before graduation. Under the counterfactual policy, all students reach level II, and therefore a student in a given school can expect the same peer quality in both level I and level II.

Table 10: Counterfactual: changing mandatory education rules

(a) Average outcomes in the population

	All		Low p.e.		Avg p.e.		High p.e.	
	(b)	(c)	(b)	(c)	(b)	(c)	(b)	(c)
graduate	0.842	0.939	0.701	0.878	0.831	0.936	0.958	0.988
enrol in high school	0.654	0.690	0.424	0.484	0.625	0.667	0.854	0.866
graduate|stay	0.967	0.939	0.930	0.878	0.964	0.936	0.992	0.988
enrol in h.s.|grad.	0.777	0.736	0.605	0.552	0.752	0.712	0.891	0.876
$C_{\text{II}}\|$ grad.	0.254	0.122	-0.381	-0.528	0.097	-0.011	0.740	0.677
$C_{\text{II}}\|$ h.s.	0.464	0.387	-0.122	-0.214	0.273	0.208	0.832	0.787

(b) Expected outcomes for three student types

	Type L_0		Type M_0		Type H_0	
	(b)	(c)	(b)	(c)	(b)	(c)
graduate	0.774	0.934	0.929	0.981	0.990	0.999
enrol in high school	0.396	0.450	0.737	0.761	0.928	0.932
graduate|stay	0.949	0.934	0.984	0.981	0.999	0.999
enrol in h.s.|grad.	0.511	0.482	0.793	0.775	0.937	0.933
$C_{\text{II}}\|$ grad.	-0.399	-0.402	0.286	0.264	0.833	0.809
$C_{\text{II}}\|$ h.s.	-0.417	-0.420	0.279	0.256	0.832	0.808

Note. Outcomes in columns (b) (i.e. "baseline") are computed using the estimated parameters of the model, while columns (c) (i.e. "counterfactual") contain results of a counterfactual simulation in which students are not allowed to drop out: they have to either graduate at $t = 2$ or stay in school until time $t = 3$. In panel a) values are the average of 1000 simulations of the structural model for each individual in the sample; the table shows the average in the sample and by subgroup of parental education (low, average, high). In panel b) values are the average of 100000 simulations for each student type with given characteristics. More specifically, L_0, M_0 and H_0 have parents with primary, upper secondary, and tertiary education respectively. Their ability is set at the average value of 0. They are assigned the mean school environments for students with low, average, and high parental background respectively. Their individual characteristics are: male, Spanish, born on July 1, began middle school at 12 years old, attends the middle school at the top of the application list. Primary school effect, cohort effect, and neighborhood quality are set at their average value.

confirmation that many early leavers would successfully graduate and in some cases even pursue further education under a different set of rules.

7 Conclusions

Suppose that a policy maker would like to identify the most "successful" schools, in order to learn their best practices and apply them in other schools. The school inputs that affect cognitive skills, as identified using the model of this study or other comparable measures of value added, would allow them to rank schools based on their ability to improve performance, as measured by a nationwide test. However, it is not evident that attending one of the top ranked schools would be of benefit to every type of student, if those schools are not committed to helping every student succeed. Indeed, while attending such schools potentially raises cognitive skills, this is not accomplished if the student drops out. Moreover, in a different school they might have graduated and acquired a stronger motivation to pursue further education, which would likely lead to better outcomes in the labor market.

The findings indicate that using school value added on performance as a proxy for "school quality" is not a benign assumption. In the best case, it might be a useful simplification in the case of students with favorable socioeconomic conditions and high ability since they are very likely to pursue further academic education regardless of their current school environment. On the other hand, school environment is a crucial determinant of educational attainment in the case of students from a less favorable socioeconomic background, not only through its contribution to cognitive skills development, but also by way of its impact on preferences. As shown in the counterfactual scenario in which dropping out is eliminated, almost all students would successfully graduate if they remain in school; however, under the current rules, a sizable proportion of them decide to drop out. Evaluating school effectiveness using only performance may lead to conclusions that do not benefit disadvantaged students: a policy maker whose goal is to improve educational outcomes for all should not ignore other dimensions, such as educational attainment.

The methodological approach proposed here makes it possible to disentangle a school's impact on attainment through its effects on skills, educational decisions, and grading policies. Future research should endeavor to open the "black box" of school inputs and understand the mechanisms that lead to the differentiation across schools. Furthermore, the results highlight the importance of studying how seemingly homogeneous rules are implemented in decentralized institutions. Thus, while the public schools in the sample share the same general set of rules for grading, retention, and graduation, differences in implementation have long-lasting effects on student outcomes.

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., P. Hull, and C. R. Walters (2022, December). Methods for measuring school effectiveness. Working Paper 30803, National Bureau of Economic Research.

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 (2020). What drives enrolment gaps in further education? the role of beliefs in sequential schooling decisions. *Economica 87*(346), 490–529.

Beuermann, D. W., C. K. Jackson, L. Navarro-Sola, and F. Pardo (2022, 06). What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output. *The Review of Economic Studies 90*(1), 65–101.

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*.

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.

Jackson, C. K., S. C. Porter, J. Q. Easton, A. Blanchard, and S. Kiguel (2020, December). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights 2*(4), 491–508.

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 2018: 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.

Appendixes

A Additional tables and figures

A.1 Tables

Table A-11: 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-12: Variation of outcomes across schools by parental education

	C_{II}	graduate	enroll in high school
Interquantile range:			
low parental education	0.410 (5.2e-04)	0.117 (1.6e-03)	0.094 (9.9e-04)
average parental education	0.391 (7.2e-04)	0.081 (1.8e-03)	0.089 (1.1e-03)
high parental education	0.417 (5.3e-04)	0.027 (9.1e-04)	0.054 (1.0e-03)
all students	0.405 (3.6e-04)	0.071 (9.0e-04)	0.078 (6.1e-04)
Standard deviation:			
low parental education	0.344 (4.7e-04)	0.077 (1.0e-03)	0.076 (7.5e-04)
average parental education	0.329 (6.3e-04)	0.054 (1.1e-03)	0.071 (8.6e-04)
high parental education	0.352 (5.0e-04)	0.020 (6.0e-04)	0.044 (7.7e-04)
all students	0.341 (3.2e-04)	0.048 (5.7e-04)	0.063 (4.7e-04)

Note. The table plots the average interquantile range (upper panel) and the average standard deviation (bottom panel) of the simulated individual outcomes in each school of the sample. Standard error of the mean in parentheses.

Table A-13: Variation of outcomes across schools by gender and immigrant status

	C_{II}	graduate	enroll in high school
Interquantile range:			
male	0.422 (4.0e-04)	0.082 (1.3e-03)	0.084 (7.9e-04)
female	0.387 (3.1e-04)	0.060 (1.3e-03)	0.072 (9.2e-04)
immigrant	0.401 (8.6e-04)	0.124 (2.1e-03)	0.099 (1.4e-03)
native	0.406 (3.9e-04)	0.062 (9.2e-04)	0.074 (6.6e-04)
Standard deviation:			
male	0.357 (3.3e-04)	0.056 (8.0e-04)	0.068 (6.0e-04)
female	0.324 (3.0e-04)	0.039 (7.9e-04)	0.057 (7.0e-04)
immigrant	0.336 (7.8e-04)	0.080 (1.3e-03)	0.077 (1.0e-03)
native	0.342 (3.5e-04)	0.042 (5.9e-04)	0.060 (5.1e-04)

Note. The table plots the average interquantile range (upper panel) and the average standard deviation (bottom panel) of the simulated individual outcomes in each school of the sample. Standard error of the mean in parentheses.

Table A-14: Variation of outcomes across schools: comparing average ability students with low and high parental background

	Student L_0				Student H_0			
	mean	p80-p20	p75-p25	s.d.	mean	p80-p20	p75-p25	s.d.
graduate	0.784	0.148	0.120	0.084	0.988	0.011	0.009	0.008
enrol in high school	0.404	0.178	0.117	0.101	0.907	0.056	0.043	0.037
$C_{\text{I},1}$	-0.309	0.543	0.442	0.367	0.632	0.543	0.442	0.367
C_{II}|grad.	-0.287	0.460	0.400	0.331	0.618	0.481	0.394	0.336
dropout at $t = 1$	0.101	0.071	0.065	0.045	0.008	0.008	0.007	0.006
drop at $t = 2$|stay	0.186	0.083	0.069	0.049	0.052	0.043	0.037	0.024
retained at t=1	0.292	0.128	0.108	0.083	0.034	0.020	0.018	0.015
retained at t=2 |stay	0.173	0.101	0.075	0.057	0.012	0.009	0.006	0.007
graduate|stay	0.940	0.062	0.050	0.034	0.998	0.002	0.002	0.001
enrol in h.s.|grad.	0.513	0.168	0.131	0.102	0.917	0.048	0.038	0.033

Note. Frequencies are constructed using 10000 simulations of the structural model for each type and every school of the sample. Type L_0 has parents with primary education; type H_0 has parents with tertiary education. They both have average ability $h = 0$, and are male, Spanish, born on July 1, they began middle school at 12 years old. Primary school effect, cohort effect, and neighborhood quality are set at their average value. The average peer characteristics in each school are used. The first column of each part of the table contains the mean outcome (weighted by school size), the other columns contain difference between quantiles and the standard deviation.

Table A-15: Educational outcomes by retention status: variation across schools for student L_0 (low parental background)

	Graduate		High school		Drop at $t = 1$		High school | grad.	
	p75-p25	s.d	p75-p25	s.d	p75-p25	s.d	p75-p25	s.d
promoted at t=1	0.067	0.045	0.127	0.101	0.050	0.034	0.125	0.102
retained at t=1	0.161	0.101	0.082	0.060	0.092	0.066	0.125	0.087
graduate at t=2	-	-	-	-	-	-	0.124	0.101
retained at t=2	0.100	0.066	0.109	0.089	-	-	0.119	0.100

Note. Statistics are computed using the simulation in Table A-14, for Type L_0 (low parental education and average ability). 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 A-16: Educational outcomes by retention status: student L_0 in his typical school environment. Baseline simulation

(a) *Ex ante* probabilities

	overall	promoted at t=1	retained at t=1
graduate	0.771	0.881	0.525
enrol in high school	0.394	0.508	0.140
retained at t=1	0.309	0.000	1.000
retained at t=2	0.099	0.143	0.000
drop. at t=1 or t=2	0.187	0.106	0.368
stay but fail at t=3	0.042	0.013	0.107

(b) Conditional probabilities

	overall	promoted	ret. at t=1	ret. at t=2
dropout at t=1	0.114	0.087	0.173	-
drop. at t=2|stay & do not grad.	0.206	-	0.236	0.129
graduate|stay	0.948	0.985	0.830	0.895
enrol in h.s.|graduate	0.511	0.605	0.267	0.375

(c) Actual and perceived skills

	overall	promoted	ret. at t=1	ret. at t=2
true $C_{\text{I},1}$	-0.416	-0.416	-0.416	-
perceived $\widehat{C}_{\text{I},1}$	-0.416	-0.388	-0.479	-
true $C_{\tau,2}$, $\tau \in \{\text{I}, \text{II}\}$	-	-0.428	-0.257	-0.428
perceived $\widehat{C}_{\tau,2}$, $\tau \in \{\text{I}, \text{II}\}$	-	-0.386	-0.281	-0.507
true C_{II}	-0.365	-0.428	-0.394	0.089
perceived $\widehat{C}_{\text{II}}$	-0.342	-0.386	-0.394	0.044

Note. Frequencies are computed using 10000 simulations of the estimated structural model for the representative student L_0. Type L_0 has parents with primary education, average ability $h = 0$, and is male, Spanish, born on July 1, they began middle school at 12 years old. Primary school effect, cohort effect, and neighborhood quality are set at their average value. Peers characteristics and school inputs take the average values among students with low parental background.

Table A-17: Educational outcomes by retention status: student L_0 in his typical school environment. Counterfactual simulation with observed h

(a) *Ex ante* probabilities

	overall	promoted at t=1	retained at t=1
graduate	0.793	0.888	0.554
enrol in high school	0.400	0.502	0.146
retained at t=1	0.286	0.000	1.000
retained at t=2	0.094	0.131	0.000
drop. at t=1 or t=2	0.169	0.101	0.340
stay but fail at t=3	0.038	0.011	0.106

(b) Conditional probabilities

	overall	promoted	ret. at t=1	ret. at t=2
dropout at t=1	0.104	0.085	0.152	-
drop. at t=2|stay & do not grad.	0.193	-	0.222	0.118
graduate|stay	0.954	0.988	0.840	0.904
enrol in h.s.|graduate	0.505	0.591	0.264	0.369

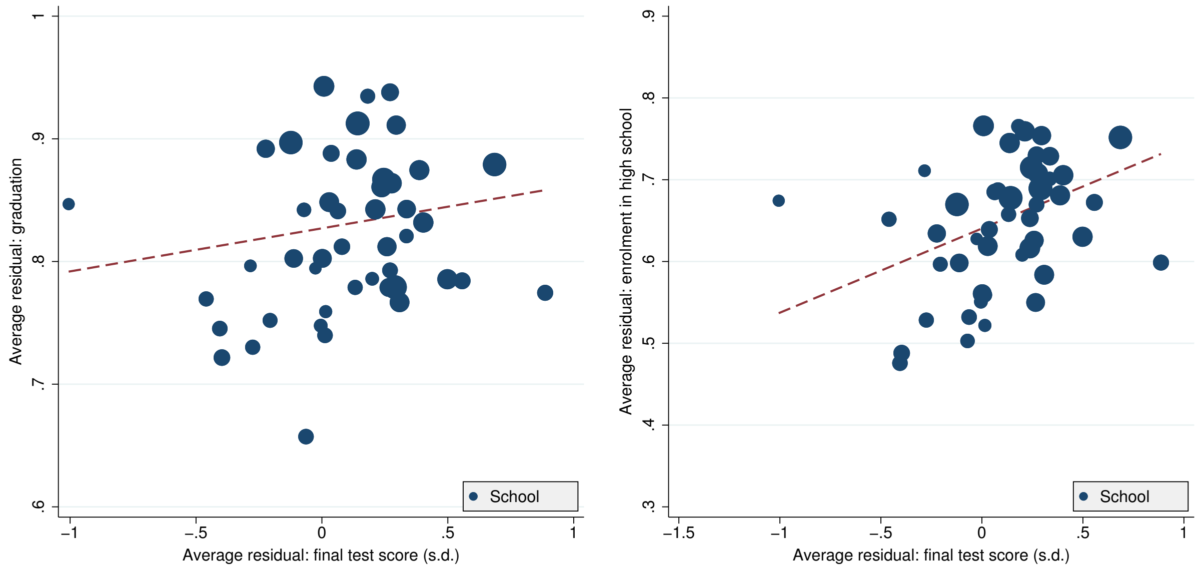
(c) Skills

	overall	promoted	ret. at t=1	ret. at t=2
true $C_{\mathrm{I},1}$	-0.416	-0.416	-0.416	-
true $C_{\tau,2}, \quad \tau \in \{\mathrm{I}, \mathrm{II}\}$	-	-0.428	-0.257	-0.428
true C_{II}	-0.369	-0.428	-0.394	0.089

Note. As in Table A-16, frequencies are computed using 10000 simulations for the representative student L_0. However, a counterfactual model with known ability is used. Student L_0 has parents with primary education, average ability $h = 0$, and is male, Spanish, born on July 1, they began middle school at 12 years old. Primary school effect, cohort effect, and neighborhood quality are set at their average value. Peers characteristics and school inputs take the average values among students with low parental background.

A.2 Figures

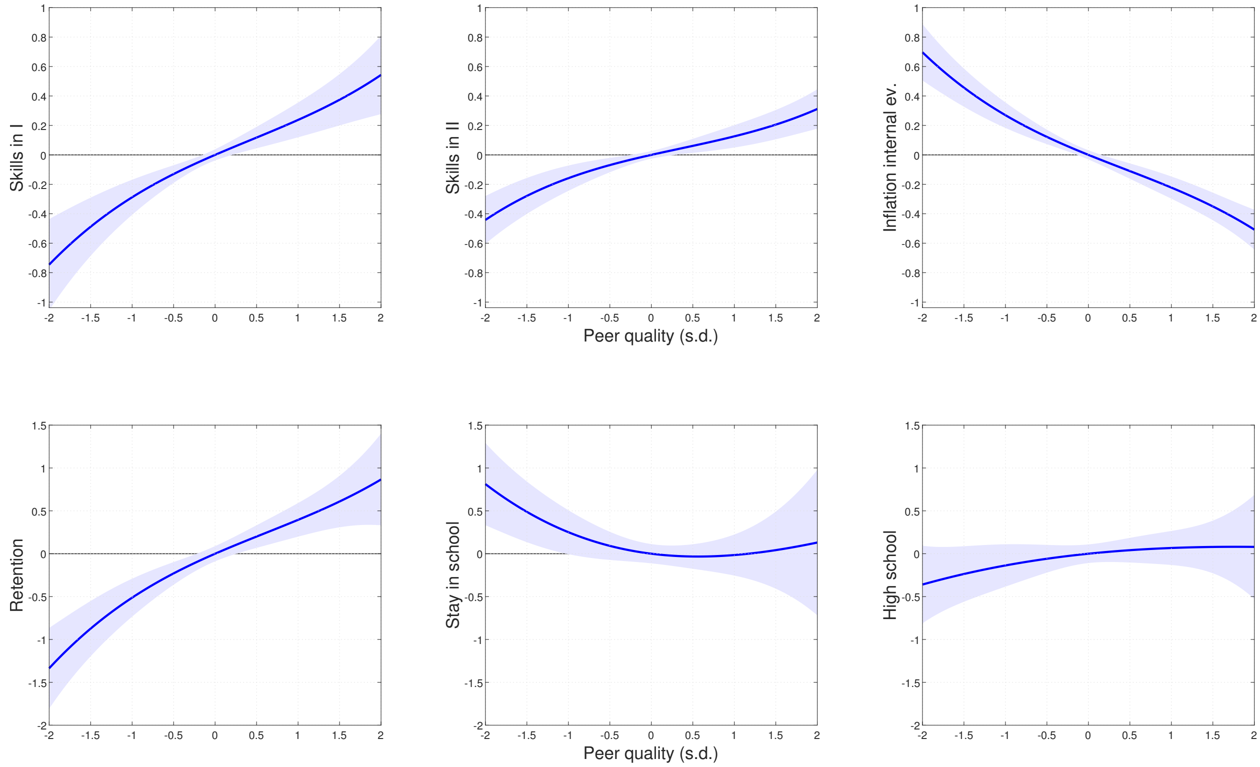
Figure A-10: Educational outcomes by school



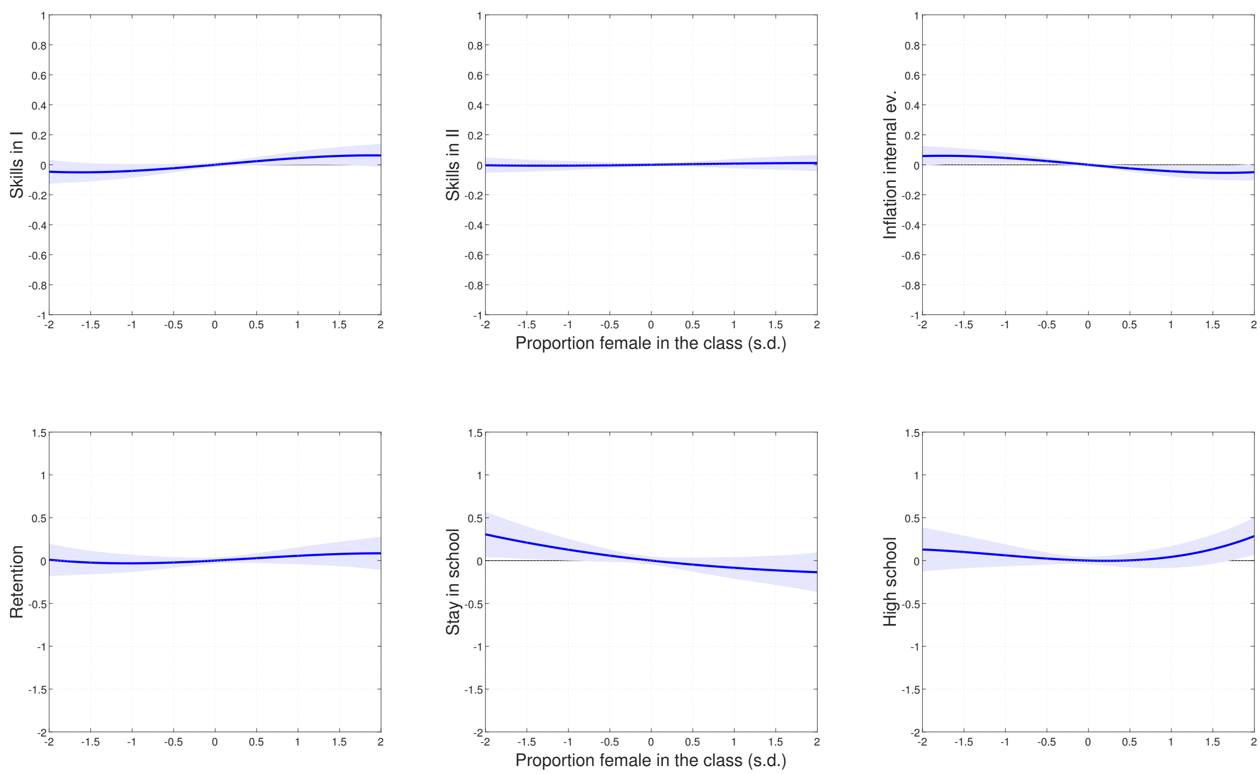
Note. Each dot plots average residuals at the school level from a regression on a vector of dummies for mother and father education (the sample mean is added back). The dot size is proportional to the school size. The dashed line is a linear fit. Correlations are 0.19 (left) and 0.34 (right); correlations weighted by school size are 0.16 and 0.39 respectively.

Figure A-11: Effects of peer variables on outcomes

(a) Peer quality

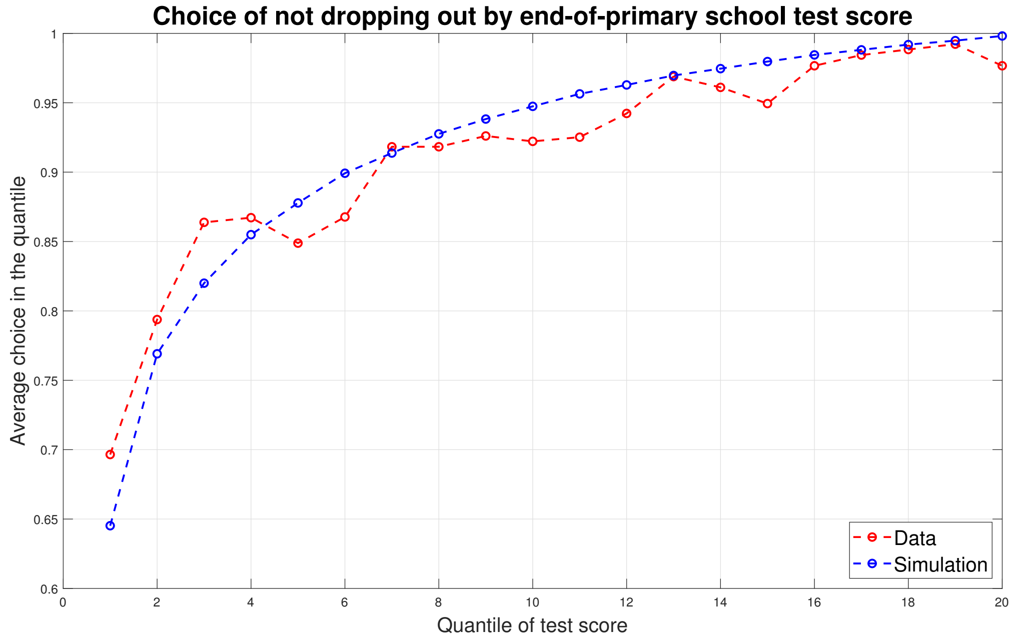


(b) Share of female



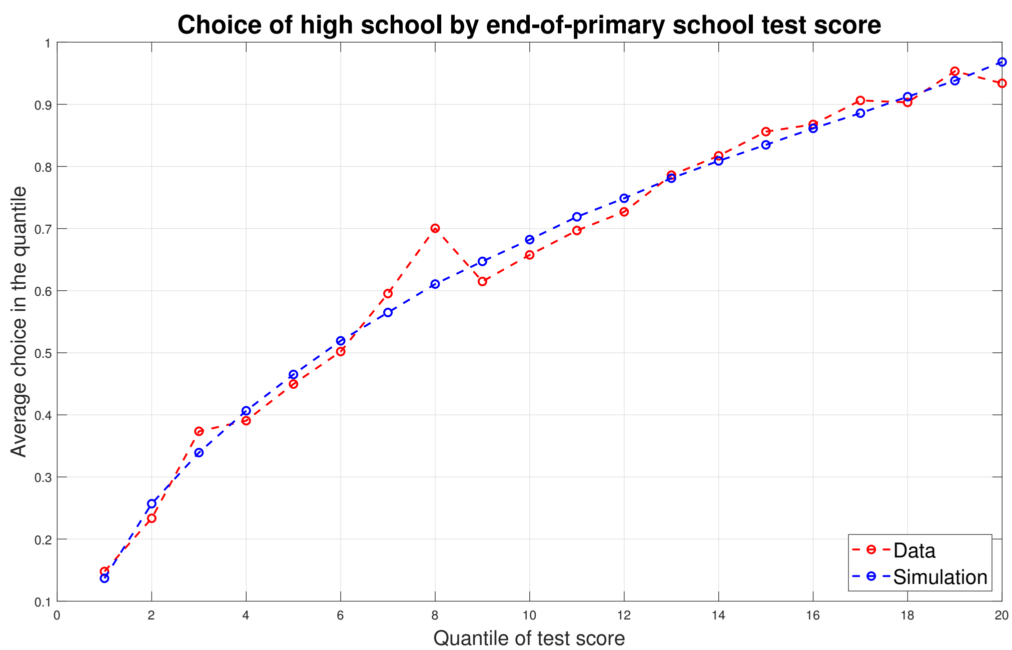
Note. The figures plot the effects of peer quality (panel (a)) and share of female (panel (b)). Shaded areas are 95% confidence intervals.

Figure A-12: Fit of the model (i)



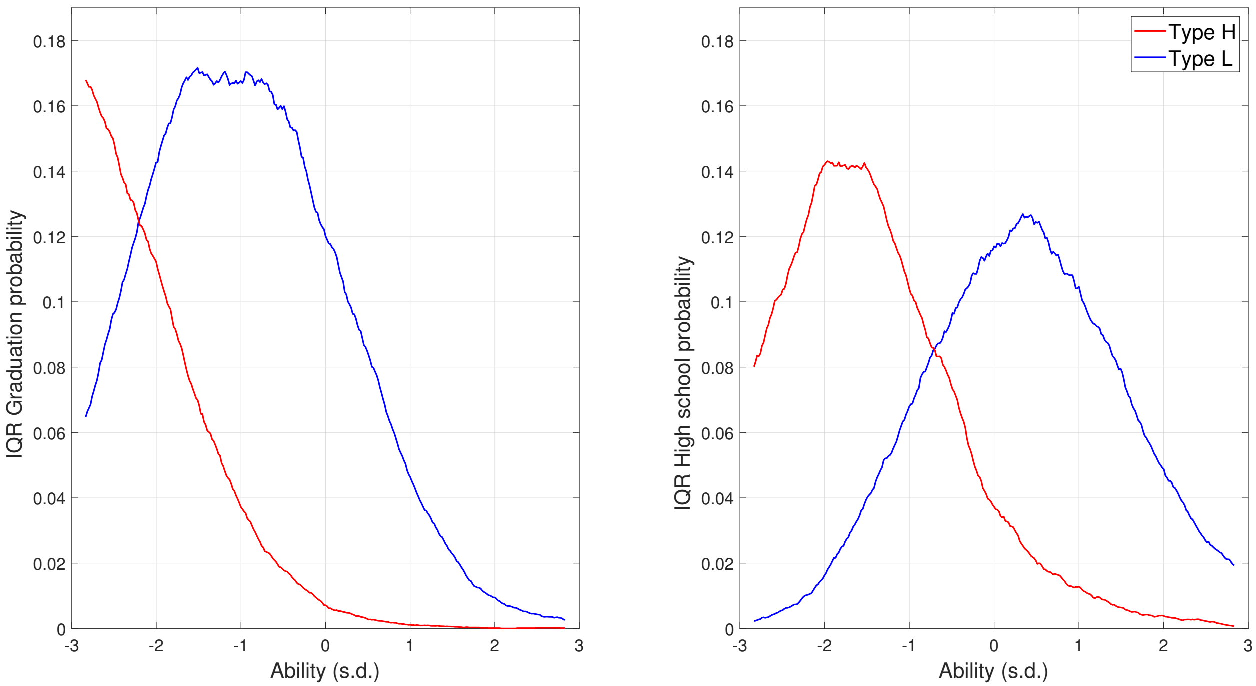
Note. The figure plots the share of students who chose to stay in school at time $t = 1$ by quantile of their test score at the end of primary school. Sample average from the real data are in red, while results of the simulation performed using the estimated parameters of the model are in blue.

Figure A-13: 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 A-14: Attainment of type L and type H students by ability



Note. The left figure plots the interquantile range (IQR)of graduation probability (y-axis) by ability level (x-axis) for representative students with highly educated Parents (red line) and low-educated parents (blue line). The IQR is the difference in graduation probability between the school at the 75th percentile and the school at the 25th percentile. Each probability is the average of 1000 simulations for a given student type at a given level of ability. For all simulations the individual characteristics are: male, Spanish, born on July 1, began middle school at 12 years old, attends the middle school at the top of the application list; primary school effect, cohort effect, neighborhood quality are set at their average values. The right figure has the same structure, but the outcome used is the probability of enrolling in high school.

B Data

B.1 Data sources

The *Departament d'Ensenyament* (the ministry of education of the autonomous community of Catalonia) provided enrollment records for all schools in Barcelona, from primary education to upper secondary education, from the school year 2009/2010 to 2015/2016. For each year, data include school and class attended by the students and information on promotion or retention.[60] In this paper I focus on public middle schools, because information on final evaluations assigned by teachers at the end of the year, and on time-invariant characteristics such as gender and nationality are available only for public schools. The data also allow me to identify children with special education needs.

A separate data source from the ministry of education contains the 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.[61] 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.[62]

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, presumably the most up-to-date information.

The previous data sources have been anonymized by the Institut Català d'Estadistica (IDESCAT), which provided unique identifiers to merge them.

Finally, I use publicly available data from the national Tax Agency (*Agencia Tributaria*) for the average income at the postal code level. I use data for year 2013, the first one for which information are available.[63]

[60] 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 next Subsection B.2 for more details.

[61] 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)

[62] 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.

[63] See `https://www.agenciatributaria.es`.

B.2 Sample creation and data cleaning

The initial 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. I drop about 2.5% of students who have special education needs. In fact, 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 the model. Furthermore, I focus the 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.

Further data cleaning includes: 1. dropping 13 students who were younger than 12 years old or older than 13 when they enroll[64]; 2. 89 students who appear in the enrollment data only once at 12/13 years old, assuming that they moved in another region; 3. 45 students for whom I could not retrieve internal evaluations in first or second grade.

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.[65]

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). The chosen thresholds exclude about 3% of the students in the selected sample. Thus, I exploit 47 out of 53 public middle schools that appear in the enrollment data in the time period under analysis.[66] The final sample consists of 1540 students.

C Estimation

This appendix provides more details about the estimation strategy summarized in Section 4. I repeat the individual likelihood here for ease of reading.

Define $d_i = (d_{it})_t$ ($t \in \{1, 2, 3\}$) to be the vector of choices of student i, $\text{fail}_i = (\text{fail}_{it})_t$ to be the vector of retention/graduation events, and $o_i = (o_{it})_t$ to be the vector of evaluations observed by i.[67] The student makes $T_d \in \{1, 2, 3\}$ decisions, receives $T_f \in \{1, 2, 3\}$ notifications of retention/graduation, and observes signals in $T_d + 1$ periods. More specifically, they receive T_d internal evaluations and $T_r \geq 1$ external evaluations. For instance, consider a student who is retained in

[64] 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.

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

[66] In a previous version of this paper, I imposed more restrictive thresholds and exploited data from only 44 schools. Overall, results are very similar.

[67] 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. I denote g_{it} (r_{it}) as the evaluation at time t, while abstracting from the level, and $g_{\tau,i}$ ($r_{\tau,i}$) as the last evaluation in level τ, while abstracting from time.

level I, stays in school one more period and then drops out. They make two choices, i.e. $d_i = (1, 0)$. At $t = 1$ they are retained, and at $t = 2$ they are promoted to level II, such that $\text{fail}_i = (1, 0)$. They are evaluated externally at the end of primary school, and internally in level I at $t = 1$ and $t = 2$, i.e. $o_i = (r_{0,i}, g_{\text{I},i1}, g_{\text{I},i2})$.

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

$$L_i = \int L(r_{i0}|h)L(\text{fail}_{i1}|h, r_{i0})L(g_{i1}|h, r_{i0})L(d_{i1}|h, r_{i0}, g_{i1})...L(d_{iT_d}|h, o_i, d_{i1}, ..., d_{it_d-1})\, \boldsymbol{\phi}(h)dh =$$
$$\Big(L(d_{i1}|r_{i0}, g_{i1})...L(d_{iT_d}|o_i, d_{i1}, ..., d_{iT_d-1})\Big) \times \Big(L(\text{fail}_{i1}|r_{i0})...L(\text{fail}_{iT_f}|o_i, d_{i1}, ..., d_{iT_d-1})\Big)\times$$
$$\times \int L(o_{iT_d}|h, d_{i1}, ..., r_{i0}, ...)...L(r_{i0}|h)\, \boldsymbol{\phi}(h)dh \tag{A-1}$$

Therefore, the log-likelihood is separable in three parts that can be estimated sequentially:

$$\log L_i = \log L_{i,d} + \log L_{i,\text{fail}} + \log L_{i,o} \tag{A-2}$$

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 C.1. 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 C.2 and C.3.

C.1 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

$$\arg\max_{\zeta,\sigma} \sum_i \log L(o_i; \zeta, \sigma) = \arg\max_{\zeta,\sigma} \sum_i \log \int L(o_i; \zeta, \sigma|h)\boldsymbol{\phi}(h)\, dh \tag{A-3}$$

The main point behind this application of the EM algorithm is that if $\widehat{\zeta}$ is a maximizer for (A-3), then it also solves

$$\arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma|h)\boldsymbol{\psi}_i(h)\, dh \tag{A-4}$$

Therefore for a given value of σ, $\widehat{\zeta}$ can be retrieved using $(A-4)$ rather than $(A-3)$. $\widehat{(\zeta, \sigma)}$ 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 $(A-4)$.

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

C.1.1 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:[68]

$$\sigma = \mathrm{E}\left(\omega_i + \mathrm{E}_i(h)\mathrm{E}_i(h)'\right), \tag{A-5}$$

The sample equivalent at step k is computed as

$$\widehat{\sigma}^k = \frac{1}{N}\left(\omega_i^k + \mathrm{E}_i^k(h)^2\right) \tag{A-6}$$

C.1.2 M-step

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

$$\arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh \tag{A-7}$$

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

$$\sum_i \int \log L(o_i; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh = \tag{A-8}$$

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

$$= \sum_i \int \log L(r_{0,i}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \sum_{it} \int \log L(g_{I,it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \tag{A-9}$$

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

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

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

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

expression can be derived as follow:[69]

$$
\begin{aligned}
-\sum_i logL_i \;=\; & \sum_i \frac{1}{2}\log(2\pi\rho_{r_0}) + \frac{1}{2\rho_{r_0}}\left(\omega_i^k + \left(r_{0,i} - (\mathrm{E}_i^k(h) + z'_{i0}\beta_0)\right)^2\right) + \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{g_\mathrm{I}}) + \frac{1}{2\rho_{g_\mathrm{I}}}\left(\omega_i^k + \left(g_{\mathrm{I},it} - (\nu_1 + \mu_\mathrm{I}\mathrm{E}_i^k(h) + z'_{it}(\beta_\mathrm{I} + \gamma) + \kappa_\mathrm{I} I_{0,i})\right)^2\right) + \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{r_\mathrm{II}}) + \frac{1}{2\rho_{r_\mathrm{II}}}\left(\omega_i^k + \left(r_{\mathrm{II},it} - (o_\mathrm{II} + \lambda_\mathrm{II}\mathrm{E}_i^k(h) + z'_{it}\beta_\mathrm{II} + \kappa_\mathrm{II} I_{\mathrm{I},i})\right)^2\right) + \\
+ & \sum_{it} \frac{1}{2}\log(2\pi\rho_{g_\mathrm{II}}) + \frac{1}{2\rho_{g_\mathrm{II}}}\left(\omega_i^k + \left(g_{\mathrm{II},it} - (\nu_\mathrm{II} + \mu\lambda_\mathrm{II}\mathrm{E}_i^k(h) + \mu(z'_{it}\beta_\mathrm{II} + \kappa_\mathrm{II} I_{\mathrm{I},i}) + z'_{it}\gamma)\right)^2\right)
\end{aligned}
\tag{A-10}
$$

The total likelihood in (A-10) 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 drop out 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 (A-10) 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 $\mathrm{E}_\epsilon(\mathrm{E}_h(\epsilon_{r_0,i}|r_{0,i}))$:

$$
\mathrm{Var}(\epsilon_{r_0,i}) = \mathrm{E}(\epsilon^2_{r_0,i}) = \mathrm{E}(\mathrm{E}((r_{0,i} - h_i - z'_{i0}\beta_0|r_{0,i})^2)) = \tag{A-11}
$$

$$
= \mathrm{E}\left(\int \left(r_{0,i} - h - z'_{i0}\beta_0\right)^2 \psi_i(h)\,dh\right) = \tag{A-12}
$$

$$
= \mathrm{E}\left[\mathrm{Var}_i(h) + \left((r_{0,i} - \mathrm{E}_i(h) - z'_{i0}\beta_0)^2\right)\right] \tag{A-13}
$$

[69] It is easy to see how to derive (A-10) from (A-9). For instance the contribution of the first region-wide test is given by:

$$
\begin{aligned}
&\int \log L(r_{0,i};\zeta,\sigma^{k-1}|\eta)\psi_i^k(h)\,dh = \int \log\left(\frac{1}{\sqrt{2\pi\rho_{r_0}}}\exp\left(-\frac{(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2}{2\rho_{r_0}}\right)\right)\psi_i^k(h)\,dh = \\
&= \int\left(-\frac{1}{2}\log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}}(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2\right)\psi_i^k(h)\,dh = \\
&= -\frac{1}{2}\log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}}\mathrm{E}_i^k\left((r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2\right) = \\
&= -\frac{1}{2}\log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}}\left(\mathrm{Var}_i^k\left(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0}\right) + \left(\mathrm{E}_i^k(r_{0,i} - h - x'_i\beta_{x,0} - p'_i\beta_{p,0})\right)^2\right) = \\
&= -\frac{1}{2}\log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}}\left(\omega_i^k + \left(r_{0,i} - \mathrm{E}_i^k(h) - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0}\right)^2\right)
\end{aligned}
$$

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

$$\rho_{r,0}^{k}=\frac{\sum_i\left(\omega_i^k+(r_{0,i}-E_i^k(h)-z'_{i0}\beta_0^k)^2\right)}{N} \tag{A-14}$$

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

C.2 Retention and graduation probabilities

I estimate the parameters of the logit model described in Section 3.3. 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 $\widehat{\mathfrak{I}}$ 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.

C.3 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 $\text{E}_{i,t}(h)$ 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

$$\begin{aligned}\text{E}_\varepsilon(u_{it}|v_{it}^A)&=\log\left(\exp(v_{it}^A)+1\right)= \\ &=\log\left(\exp(\phi_{A,r}\text{E}_{i,t}(C_{\text{II}})+y'_{it}\theta_A)+1\right)\end{aligned} \tag{A-15}$$

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:

$$\mathrm{E}_{i,t}(h) = (\widehat{h}_{it-1}, \mathbf{s}'_t)\mathbf{c}_t = c_{0,t}\widehat{h}_{it-1} + c_{1,t}s_{r,t} + c_{2,t}s_{g,t} \tag{A-16}$$

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

$$\begin{aligned}\mathrm{E}_{i,t-1}(u_{it}|\mathrm{grad}_{it} = 1) &= \int \log\left(\exp(v_{i,A}) + 1\right) \cdot \widehat{\psi}_{it-1}(s_{g,t}, s_{r,t})\, d\mathbf{s}_t = \\ &= \int \log\Big[\exp\left(\phi_{A,r}\left(k_{\mathrm{II}}I_{t-1} + z'_{it}\beta_{\mathrm{II}}\right) + y'_{it}\theta_A\right)\cdot \exp\Big(\phi_{A,r}c_{0,t}\widehat{h}_{it-1}\Big)\cdot \\ &\cdot \exp\left(\phi_{A,r}(c_{1,t}s_{r,t} + c_{2,t}s_{g,t})\right) + 1\Big] \cdot \widehat{\psi}_{it-1}(s_{g,t}, s_{r,t})\, d\mathbf{s}_t\end{aligned} \tag{A-17}$$

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^M_{i2}:

$$v^M_{i2}(\widehat{h}_{i2}, z_{i2}) + \varepsilon_{i2} = U^M_{i2} + \delta\widehat{\Pr}(\mathrm{grad}_{i3} = 1)\int \log\Big(\exp(v^A_{i3}) + 1\Big) \cdot \widehat{\psi}_{i2}(s_{g,3}, s_{r,3})\, d\mathbf{s}_3 \tag{A-18}$$

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, $\mathrm{E}_{\varepsilon}(u_{i2}|v^M_{i2}) = \log\Big(\exp(v^M_{i2}) + 1\Big)$ and $E_{i,1}(u_{i2}|\mathrm{grad}_{i2} = 0) = \int \log\Big(\exp(v^M_{i2}) + 1\Big) \cdot \widehat{\psi}_{i,1}(\mathbf{s}_1)\, d\mathbf{s}_1$. Finally, we can compute values for the first period:

$$v^M_{i,1}|(\mathrm{fail} = 1) + \varepsilon_{i,1} = U^M_{i,1} + \delta E_{i,1}(u_{i2}|\mathrm{grad}_{i2} = 0) \tag{A-19}$$

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

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

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

$$p_{it}(d_{it} = 1|d_{it-1} = 1) = \frac{\exp(v_{it})}{1 + \exp(v_{it})} \tag{A-21}$$

I maximize the total loglikelihood to estimate the parameters, following Rust (1987).[70]

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

C.4 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.

D 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

$$\log L(o_i;\zeta,\sigma) = \log \int L(o_i;\zeta,\sigma|h)\phi(h)\,dh \tag{A-22}$$

$$L(o_i;\zeta,\sigma|h) = L(r_{i,0};\zeta,\sigma|h)L(g_{i,1};\zeta,\sigma|h)...L(o_{i,T_d};\zeta,\sigma|h) \tag{A-23}$$

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

$$L(r_{i0};\zeta,\sigma|h) = \frac{1}{\sqrt{2\pi\rho_{r_0}}}\exp\left(-\frac{(r_{i,0}-h-z_{i,0}'\beta_0)^2}{2\rho_{r_0}}\right) \tag{A-24}$$

Taking the log of (A-23) 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:

$$\frac{\partial}{\partial\zeta}\sum_i \log L(o_i;\zeta,\sigma) = \sum_i \frac{1}{L(o_i;\zeta,\sigma)}\int \frac{\partial L(o_i;\zeta,\sigma|h)}{\partial\zeta}\phi(h)\,dh = 0 \tag{A-25}$$

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

$$\psi_i(h) = \frac{L(o_i;\zeta,\sigma|h)\phi(h)}{L(o_i;\zeta,\sigma)} \tag{A-26}$$

Now, moving $L(o_i;\zeta,\sigma)$ under the integral and multiplying by $1 = \frac{L(o_i;\zeta,\sigma|h)}{L(o_i;\zeta,\sigma|h)}$, equation (A-25) can be rewritten as

$$\sum_i \int \frac{L(o_i;\zeta,\sigma|h)\phi(h)}{L(o_i;\zeta,\sigma)} \frac{1}{L(o_i;\zeta,\sigma|h)} \frac{\partial L(o_i;\zeta,\sigma|h)}{\partial \zeta} \, dh = \tag{A-27}$$

$$= \sum_i \int \frac{1}{L(o_i;\zeta,\sigma|h)} \frac{\partial L(o_i;\zeta,\sigma|h)}{\partial \zeta} \boldsymbol{\psi}_i(h) \, dh = \sum_i \int \frac{\partial}{\partial \zeta} \Big(\log L(o_i;\zeta,\sigma|h) \Big) \boldsymbol{\psi}_i(h) \, dh = \tag{A-28}$$

$$= \frac{\partial}{\partial \zeta} \left[\sum_i \int \log L(o_i;\zeta,\sigma|h) \boldsymbol{\psi}_i(h) \, dh \right] = 0 \tag{A-29}$$

Thus if $\widehat{\zeta}$ solves equation (A-25) it solves also equation (A-29) 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

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

The general theory ensures convergence of the algorithm.[71]

E 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-18 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 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.[72] 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

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

[72] 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.

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. Overall, results are qualitatively and quantitatively very similar. 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.

Table A-18: 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.

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it