Universitat Autònoma de Barcelona

ESSAYS IN ECONOMICS OF EDUCATION

ANNALISA LOVIGLIO

Essays in Economics of Education

Author: Annalisa Loviglio

Supervisors: Caterina Calsamiglia

Joan Llull

May 5, 2019

A dissertation submitted to the *Departament d'Economia i d'Història Econòmica* at the *Universitat Autònoma de Barcelona* in fulfillment of the requirements for the degree of Doctor of Philosophy by the International Doctorate in Economic Analysis (IDEA).

To mom and dad

Acknowledgements

First and foremost, I am extremely grateful to my supervisors Caterina Calsamiglia and Joan Llull, for their guidance, encouragement, and continuous support. They set the example of the academic I would love to become: someone who is passionate about research, but who also values teaching, and who always finds the time to help colleagues and students.

I would also like to thank Peter Arcidiacono, Manuel Arellano, and Nicola Pavoni for hosting me at Duke University, Cemfi, and Università Bocconi, and for many useful conversations and valuable insights.

I gratefully 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.

I am indebted to the entire faculty of the IDEA program. A group of dedicated professors made the coursework period a challenging, but truly rewarding learning experience. Their availability and their thoughtful feedback have been essential for the development of my work throughout the PhD. This year, I would not have made it through the job market process without their help in perfecting my presentation skills and their moral support. A big thank you also to Àngels and Mercè for their constant assistance and their kindness.

My sincere thanks to all my current and previous colleagues: they have been essential for my professional and personal growth during the time we spent together. I learned and shared so much, from and with them. Thank you to the amazing first year TAs, and to the classmates who spent countless hours with me to find the most satisfactory solution to a difficult problem. To those I taught in the following years, when I truly understood that teaching and learning are two sides of the same coin. To many colleagues, especially those in my cohort, for the passionate conversations about research and everything else, for being cooperative rather than competitive, for sharing ideas and fears, achievements and failures. To my fellow job market candidates, for supporting each other and sharing some good laughs during a quite peculiar period of our life.

Finally, I would like to thank my friends, near and far, and my family, for their love and encouragement during these years.

Preface

This thesis studies how the education system in place affects the human capital development of its students. Chapter 1 explores the role of schools, Chapter 2 studies the grading system, and Chapter 3 investigates the consequence of a specific regulation requiring that all students start primary education in the calendar year in which they turn 6.

For the empirical analyses, I gathered and studied administrative data on attainment, test scores, socio-economic and demographic characteristics of the universe of Catalan students enrolled in primary and secondary education from 2009 to 2015. Chapter 1 focuses on public schools in Barcelona, while Chapter 2 and Chapter 3 exploit data for the entire region. The second and third chapters are coauthored with Caterina Calsamiglia.

In Chapter 1, I study how the school environment affects students' cognitive skills and educational attainment. I estimate a dynamic structural model of cognitive skills accumulation and educational decisions of students enrolled in lower secondary education, using data for the universe of public schools in Barcelona. Its key feature is that it allows me to separately identify the different channels through which schools affect student outcomes. I find large variation across schools both in their effect on cognitive skills development, and in their effects on students' educational choices above and beyond their level of cognitive skills. School environment is particularly relevant for choices of students with disadvantaged family background. Moreover their probabilities of graduating or enrolling in upper secondary education if they attend a given middle school have limited correlation with their expected performance in that school. Results suggest that evaluating and comparing schools using only nation-wide assessments may not favor disadvantaged students, who particularly benefit from schools which increase educational attainment, not only test scores.

In Chapter 2, we study the differences between the evaluations assigned by teachers (GPA) and results in region-wide tests. We show that the GPA is strongly deflated in classes of above-average students. In other words, having better peers harms the evaluation obtained by a given student. Student access to education levels, tracks or majors is usually determined by their previous performance, measured either by internal exams, designed and graded by teachers in school, or external exams, designed and graded by central authorities. Our findings put forth a source of distortion that may arise in any system that uses internal grades to compare students across schools and classes. We also find suggestive evidence that school choice is impacted only the year when internal grades matter for future prospects.

In Chapter 3, we study the effect of students' age at enrollment in primary school on their educational outcomes throughout primary and secondary education. Having a unique cut-off to determine when children can access school induces a large heterogeneity in maturity to coexist in a classroom. We show that relatively younger children do significantly worse both in tests administered at the school level and at the regional level, and they experience greater retention. These effects are homogeneous across socioeconomics and significant across the whole distribution of performance. Moreover younger children exhibit higher dropout rates and chose the academic track in secondary school less often. Younger children are also more frequently diagnosed with learning disabilities.

Contents

List of Tables

List of Figures

Chapter 1

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

1.1 Introduction

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

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

[2] The "No Child Left Behind" act (replaced by the "Every Student Succeeds" act in 2015) requires public schools to administer a statewide standardized test annually; if school's results are repeatedly poor

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

A large literature shows that life success depends on more than cognitive skills alone, and that interventions aimed at raising a broader set of skills have impressive returns in the long run, contributing to bridge the gaps due to family conditions.[3] These results emphasize that the debate around school quality should go beyond test scores alone. In fact, a child is left behind not only if she gets a low score in a standardized test, but also if she is not provided with the appropriate school environment to develop both her cognitive and non-cognitive skills, and to motivate her to pursue further studies. Secondary education is a crucial stage, because for the first time in their educational career students can choose whether they want to acquire further education and, in some cases, what they want to study. In fact, in most countries basic education is compulsory, but students are legally allowed to leave when they reach a given age, not upon completion of a given level. Moreover, in many European countries after completing lower secondary education students choose whether to enroll in the track which gives access to University. This decision is typically taken when they are 16 years old or even younger.[4] A student at risk of dropping out may be better off attending a school in which she feels comfortable and she is able to achieve a diploma, rather than another institution that would have potentially raised her final test score more, but from where she would have dropped out. Similarly, the choice of undertaking upper secondary or tertiary education may depend on previous performance, but also on student's motivation or family support.

In this paper, I estimate a dynamic model of cognitive skills accumulation and schooling decisions throughout lower secondary education of students enrolled in heterogeneous middle schools. At each time, cognitive skills growth depends on an unobserved ability,

---

various steps are taken to improve the school. Since 1992, U.K. has been has published so-called "school league tables" summarizing the average GCSE results in state-funded secondary schools. Underperforming schools face various sanctions (Leckie and Goldstein, 2017).

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

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

individual characteristics, and school environment (captured by classmates' characteristics and school effects). The student has imperfect information on her level of cognitive skills, but progressively learn about her true ability through various assessments. After updating her beliefs, she chooses whether to pursue further education. Her flow utility depends on her beliefs about cognitive skills, but also on the school environment and on individual characteristics. Importantly, before taking the decision, the student can be retained (i.e. required to repeat a level to stay in school). Retention may raise performance in the following period, but it increases the time needed to graduate and can change students' preferences.

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

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

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

Estimation results show that the school environment is an important determinant of cognitive skills, both through peer effects (being with 1 standard deviation higher ability peers increases cognitive skills of more than 0.1 standard deviation) and through the school effect beyond peers (the interquantile range of school effect is about 0.3 s.d.). Moreover the school environment has a sizable direct effect on educational choices (being in a school at the 75 percentile of the distribution of school effects on the choice rather than at the 25 percentile increases flow utility from enrolling in high school as much as an increase of 0.6 s.d. in cognitive skills). Results show that about half of the total variance of cognitive skills is due to unobserved ability. Evaluations are informative and students quickly acquire accurate beliefs.

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

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

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

This paper relates to several strands of the literature, particularly the literature on school quality, on human capital development, and on decision making. School accountability requires to develop reliable measures of school quality to compare among them schools (Allen and Burgess, 2013; Angrist, Hull, Pathak, and Walters, 2017; Kane and Staiger, 2002), or school types (e.g. charter *versus* traditional public schools, as in Dobbie and Fryer (2011) or Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011)). This is typically done with a value added approach, i.e. the estimation of the net effect of attending a given institution on a relevant outcome.[5] Test scores have been the most used outcome to capture "quality", under the assumptions that performances in school measures cognitive skills and are positively correlated with desirable outcomes in subsequent educational stages and in the labor market. This paper has a comparable "value added" approach, but it exploits a variety of outcomes, showing that school effects on performance and attainment are not aligned.

In fact, other lines of the literature have well established that cognitive skills alone do not explain educational choices and attainments which matter for all future life outcomes. On one hand, literature on human capital development has fully acknowledged that returns from non-cognitive skills are as high as the one from cognitive skills and the former may not be well captured using test scores (e.g. Heckman and Rubinstein (2001)). On the other hand, literature on decision making applied to educational choices shows that there is large variation among individuals with identical prior performances, due to a multitude of reasons, from differences in beliefs on the return of each choice, to different consumption values.[6] These works mainly focus on individual traits and preferences, and on their differences by gender and socio-economic background.[7] This paper contributes to

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

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

[7]An interesting example is given by Belfield, Boneva, Rauh, and Shaw (2018). They collect survey data on students' motives to pursue upper secondary and tertiary education in UK. They find that beliefs

incorporate their finding in the study of school quality. In fact it seems very plausible that school environment, similarly to the family environment, may substantially contribute to non-cogntive skills development, tastes formation, and provision of information on returns from education.[8] Having a structural model of cognitive skills development, retention, and choices allows me to first estimate the effect of school environment on cognitive skills, and then its direct effect on a given choice on top of the impact that goes through the effect on cognitive skills.

This is also what differentiates my work from other recent contributions (Angrist, Cohodes, Dynarski, Pathak, and Walters, 2016; Deming, Hastings, Kane, and Staiger, 2014) which exploit across schools differences in educational choices and attainment, such as high school graduation, college enrollment, college persistence. In fact they use those outcomes as measures of human capital alternative to test scores, for instance to show that the large gain in attending some charter schools found by previous works is not due to a "teaching to the test" attitude, but to a true improvement of skills that matter in the long run. Their analyses cannot assess whether improvements in graduation rate or college enrollment are due to the improvement of cognitive skills as measured by standardized tests, or to other factors on top of that. The advantage of the structural model implemented in this paper is that it allows to disentangle school effect through cognitive skills and through other channels, and to assess how the two aspects interact.

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

Finally, my work contributes to the debate on the effectiveness of grade retention. Despite retention being a common practice in many countries, empirical literature provides mixed evidence of its effectiveness in improving student performances (Allen, Chen, Willson, and Hughes, 2009; Fruehwirth, Navarro, and Takahashi, 2016). My results suggest that it can improve test scores at the end of middle school (at the cost of longer time in education); however it has a negative effect of students' consumption value of schooling, therefore the net result is a large increase in dropout rate among retained students and

about future consumption values play a more important role than beliefs about the monetary benefits and costs, and differences in the perceived consumption value across gender and socio-economic groups can account for a sizeable proportion of the gender and socio-economic gaps in students' intentions to pursue further education.

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

lower probability of enrollment in high school.[9] Interestingly, the gap is larger for students of relatively higher ability, who would not be at risk of retention in schools that are more lenient.

The remainder of this paper is organized as follows. Section 2.3 provides background on the Spanish education system, describes the data, and discusses descriptive statistics. Section 2.2 describes the model and Section 1.4 details the estimation procedure. Section 2.4.1 presents the estimation results. Section 1.6 studies how the school environment affect educational outcomes of children with low or high parental background. Section 1.7 studies how raising school effects change average outcomes in the entire population and among students with low parental background. Section 1.8 concludes.

1.2 Data

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

1.2.1 Education system

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

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

After successfully completing lower secondary education, students can enroll in high

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

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

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

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

1.2.2 Data sources and sample selection

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

The *Consell d'Avaluació de Catalunya* (public agency in charge of evaluating the educational system) provided me with the results of standardized tests taken by all the students in the region attending 6th grade of primary school and 4th grade of middle school. Such tests are administered in the spring since 2008/2009 for primary school and since 2011/2012 for middle school. They assess students' competence in Maths, Catalan, and Spanish.[12] These exams are externally designed and graded. In this paper I refer

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

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

[12]The tests are low stakes, because they do not have a direct impact on student evaluations or progress

to the test scores as *external evaluations*, in contrast with the final evaluations given by teachers in the school, which I call *internal evaluations*. The tests are administered in two consecutive days in the same premises in which students typically attend lectures. Normally every student is required to take all the tests, however children that are sick one or both days and do not show up at school are not evaluated.[13] In the analysis I use z-scores for both internal and external evaluations. I focus my analysis on students for whom I could retrieve the evaluations in primary school, i.e. 85% of the students who enroll in a public middle school in the period under analysis.[14]

Finally, information on the student's family background, more specifically on parental education, are collected from the Census (2002) and local register data (*Padró*). When the information can be retrieved from both sources, I impute the highest level of education, presumably the most up-to-date information. I allow for three level of education: Low (at most lower secondary education), Average (upper secondary education), High (tertiary education).

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

1.2.3 Descriptive statistics

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

As shown in Table 1.1, there are wide differences across subgroups of the population. Children with low educated parents have much lower test scores when they start middle school, they are more likely to dropout at 16, and only 70% of them complete lower secondary education (the share is 95% among children with highly educated parents). Those who manage to graduate have on average lower test scores (-0.32 versus 0.59 for students with high parental background) and are less likely to pursue further studies.

to the next grades but they are transmitted to the principal of the school, who forwards them to the teachers, families and students. More information can be found at: http://csda.gencat.cat/ca/arees_d_actuacio/avaluacions-consell (in Catalan)

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

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

Only 42% of the initial pool of students with low educated parents enroll in high school, while 86% of students with highly educated parents do so.

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

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

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

The last four rows of Table 1.1 show descriptive statistics by retention status. Students are grouped into four categories: those who are never retained before leaving middle school, those who were retained in primary school (8% of the sample), those who are retained for the first time in middle school before reaching the last grade (15%), those who are retained for the first time in the last grade (4%). Students who were already behind before turning 16 years old are significantly more likely to be early dropout, especially if they have been retained during secondary education: 30% of them immediately leave schools, while only 3% of students with a regular progression dropout. Moreover less than half of them graduate and very few enroll in high school. Students who repeat the last grade are less likely to graduate and enroll in high school than the average but they have better odds than students retained at an early stage.

Table 1.2 describes the distribution of incoming students' characteristics and their outcomes at the school level; each column shows values of a given variable at various quantiles. It confirms that schools are quite different both in the types of students they teach and in the outcomes they produce. For instance, the school at the 75th percentile of average incoming test scores is more than 0.5 s.d. better than the school at the 25th percentile; the interquantile range of the share of students with low parental education is 19 p.p; in the school at the 90th percentile 81% of students enroll in high school, while only 44% do so in the school at the 10th percentile.

1.3 Model

1.3.1 Overview

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

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

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

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

1.3.2 Time line

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

- At the end of primary education, each student undertakes a nation-wide test. After completing primary education, they are assigned to a middle school and begin lower secondary education (time $t = 0$).

- Lower secondary education covers two levels (I and II). The normal length of a level is one time period, but students may be retained once, either during level I or during level II; in this case if they do not leave education they have to spend one more period in the same level.

- At time $t = 1$ students finish their first period in school; they receive internal grades and the communication of whether they have been promoted to level II. From next period education is not compulsory anymore, thus they have to decide whether to stay in school or dropout.

  - Retained students who continue in school repeat level I. At time $t = 2$ they receive new internal evaluations and they are surely promoted to level II.
  - Promoted students who continue in school access level II. At time $t = 2$ they receive internal evaluations and external evaluations from a nation-wide test, moreover they are informed of whether they successfully complete lower secondary education or they have been retained.

- At time $t = 2$ students who did not complete lower secondary education yet decide whether to leave school or to stay in level II. If they stay, at time $t = 3$ they receive internal and external grades; moreover they are told whether they graduate or not.
- Students who successfully complete lower secondary education (either at $t = 2$ or at $t = 3$) decide whether to enroll in high school to undertake further academic education.

1.3.3 Cognitive skills formation

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

The contemporaneous inputs are the unobserved ability h_i and a set of covariates $(x_i', rep_{it}, p_{it}', s_{it})'$, which include a vector of time-invariant individual characteristics $x_i = (x_{i1}, ..., x_{iJ})'$ (e.g. gender, nationality, parents' education), a dummy rep_{it} which takes value 1 if the level is repeated for the second time, a vector of peers characteristics $p_{it} = (p_{it1}, ..., p_{itK})'$ (e.g. average parental education of peers), and school effects s_{it}. Peers at time t are the other students attending the same class; class composition may change overtime because sometime classes are shuffled at the beginning of a new school year, and because some students dropout.[15]

$$C_{i,0} = z_0'\beta_0 + h_i \tag{1.1}$$

$$C_{i,\mathrm{I}t} = \alpha_{\mathrm{I}} C_{i,0} + z_{it}'\beta_{\mathrm{I}} + \mu_{\mathrm{I}} h_i, \qquad t \in \{1, 2\} \tag{1.2}$$

$$C_{i,\mathrm{II}t} = \alpha_{\mathrm{II}} C_{i,\mathrm{I}} + z_{it}'\beta_{\mathrm{II}} + \mu_{\mathrm{II}} h_i, \qquad t \in \{2, 3\}. \tag{1.3}$$

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

The notation $z'_{it}\beta_\tau$ is used for simplicity, from time $t = 1$ the functional form includes interaction between school effects and individual time-invariant characteristics. More specifically, for $t \geq 1$:

$$z'_{it}\beta_\tau = \sum_{i=1}^{J} \beta_{\tau,x_j} x_{ij} + s_i + \sum_{i=1}^{J} \beta_{\tau,sx_j}(s_i x_{ij}) + \beta_{\tau,rep}\, rep_{it} + \sum_{i=1}^{K} \beta_{\tau,p_k} p_{itk} \tag{1.4}$$

This specification allows schools to affect differently students with differing characteristics (e.g. different family background) while remaining parsimonious on the the number of parameters to estimate.[16]

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

1.3.4 Evaluations as signals

External evaluations

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

$$r_{\tau,it} = o_\tau + \lambda_\tau C_{i,\tau t} + \epsilon_{r_\tau,it}, \qquad \epsilon_{r_\tau,it} \sim \mathcal{N}(0, \rho^r_\tau). \tag{1.5}$$

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

$$r_{0,i} = C_{i,0} + \epsilon_{r_0}, \qquad \epsilon_{r_0,i} \sim \mathcal{N}(0, \rho^r_0), \tag{1.6}$$

and those who stay in school long enough also receive

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + \lambda_{\mathrm{II}} C_{i,\mathrm{II}t} + \epsilon_{r_{\mathrm{II}},it}, \qquad \epsilon_{r_{\mathrm{II}},it} \sim \mathcal{N}(0, \rho^r_{\mathrm{II}}), \tag{1.7}$$

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

Internal evaluations

At the end of each period in secondary education, students receive teachers' evaluations. Given that exams are designed and graded internally, teachers' biases or comparison with peers may affect the assigned score. Moreover schools may be more or less lenient, and

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

administer more or less difficult tests. In other words, children with the same level of underlying cognitive skills may expect to receive different evaluations depending on their characteristics, peers, or school in which they are enrolled. Moreover, similarly to nation-wide test scores, there is an exogenous normally distributed error:

$$g_{\tau,it} = \nu_\tau + \mu_\tau C_{i,\tau t} + z'_{it}\gamma_\tau + \epsilon_{g_\tau,it}, \qquad \epsilon_{g_\tau,it} \sim \mathcal{N}(0, \rho^g_\tau). \tag{1.8}$$

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

Identification of the grade equations

The scale factors in the grade equations, the shares α_τ, and the coefficients β_τ cannot be identified separately. Therefore, I will not be able to disentangle the contemporary effect of time invariant characteristics, but only their cumulative effect. Moreover a necessary assumption for identification is that school and teachers' policy for grading is constant across levels, i.e. $\gamma_{\mathrm{I}} = \gamma_{\mathrm{II}} = \gamma$.[17]
With some abuse of notation, I redefine evaluations as follow:

$$r_{0,i} = z'_{i0}\beta_0 + h_i + \epsilon_{r_0,i} \tag{1.9}$$

$$g_{\mathrm{I},it} = \nu_{\mathrm{I}} + z'_{it}(\beta_{\mathrm{I}} + \gamma) + \kappa_{\mathrm{I}} I_{0,i} + \mu_{\mathrm{I}} h_i + \epsilon_{g_{\mathrm{I}},it} \tag{1.10}$$

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + z'_{it}\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i} + \lambda_{\mathrm{II}} h_i + \epsilon_{r_{\mathrm{II}},it} \tag{1.11}$$

$$g_{\mathrm{II},it} = \nu_{\mathrm{II}} + \mu(z'_{it}\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i} + \lambda_{\mathrm{II}} h_i) + z'_{it}\gamma + \epsilon_{g_{\mathrm{II}},it}, \tag{1.12}$$

where $I_{\tau-1,i}$ is the portion of previous cognitive skills that comes from time-varying observed covariates.[18] The coefficients in β_τ capture the cumulative effects of time invariant regressors, and the innovation of time-varying regressors. Moreover, $\mu_{\mathrm{II}} = \mu\lambda_{\mathrm{II}}$.

Signals

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

[17] The latter assumption is necessary because external evaluations are observed only in level II.
[18] For instance, using the previous notation, for a student who did not repeat first level: $I_{i,\mathrm{I}} = p_{i\mathrm{I}}\beta_{p\mathrm{I}} + \alpha_{\mathrm{I}} s_{i0}\beta_{s0}$

update their beliefs on their level of cognitive skills. More specifically,

$$s(r_{\tau,it}) = h_i + \frac{1}{\lambda_\tau}\epsilon_{r_\tau,it} \tag{1.13}$$

$$s(g_{\tau,it}) = h_i + \frac{1}{\mu_\tau}\epsilon_{g_\tau,it}. \tag{1.14}$$

All students observe $r_{0,i}$ and $g_{\text{I},it}$, while the other signals they receive depend on their choices and on whether they are retained. After receiving one or more signals, students can compute the posterior distribution of their ability. When a new signal arrives, one can update the posterior distribution using the previous posterior as prior.[19]

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

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

1.3.5 Retention and graduation

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

$$\Pr(\text{failI}_i = 1|w_{i1}) = \frac{\exp(w'_{i1}\zeta_{\text{I}})}{1 + \exp(w'_{i1}\zeta_{\text{I}})}. \tag{1.15}$$

The set w_{i1} includes time invariant individual and peer characteristics, initial peers, middle school dummies, and prior beliefs $\text{E}_{i,0}(C_I)$. This specification accounts for the fact that there is no deterministic rule in place to determine retention, in particular schools can choose to be more or less lenient. While I allow prior beliefs to enter the retention probability, I do not allow C_{iI} or equivalently h_i itself to enter the equation. This would make the model very difficult to treat, because students could learn about their ability

[19] See DeGroot (1970)

through the realization of the event.[20] Moreover, this assumption appears sensible because the school personnel do not know either the true h_i when deciding about retention, but they can form a belief about it, exactly as the student does.

Individual are at risk of graduation only if they are in level II. For students who did not repeat period I the first failure is equivalent to a retention. Similarly to definition (1.15), I assume that the conditional probability takes a logit form:

$$\Pr(\text{grad}_{it} = 1 | w_{it}) = \frac{\exp(w'_{it}\zeta_{\text{II}})}{1 + \exp(w'_{it}\zeta_{\text{II}})}, \qquad t \in \{2, 3\}. \tag{1.16}$$

w_{it} includes individual and peer characteristics, middle school dummies, and prior beliefs $\text{E}_{i,t-1}(C_{II})$. $Pr(\text{grad}_{i2} = 1|w_{i2})$ can be set to 0 for individuals who are not in level II at time 2, i.e. who were retained in period I. For ease of notation in next section I will use the expression $(\text{grad}_{i2} = 0)$ also for those with $\text{failI}_i = 1$.
Students are assumed to know the parameters and form expectations over their probability of graduation using (1.16).

1.3.6 Flow utilities

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

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

$$\begin{aligned} U^M_{it} &= \phi_M \text{E}_{i,t}(C_\tau) + \text{ret}'_{M,it}\theta_{M,r} + x'_i\theta_{M,x} + p'_{it}\theta_{M,p} + s'_{it}\theta_{M,s} + \quad (1.17)\\ &+ \text{E}_{i,t}(C_\tau)\Big[\text{ret}'_{M,it}\kappa_{M,r} + \kappa_{M,x}(x'_i\theta_{M,x}) + \kappa_{M,p}(p'_i\theta_{M,p}) + \kappa_{M,s}(s'_{it}\theta_{M,s})\Big] + \varepsilon_{it} = \\ &= (\phi_M + \widetilde{y}'_{it}\kappa_M)\text{E}_{i,t}(C_\tau) + y'_{it}\theta_M + \varepsilon_{it}, \quad (1.18)\end{aligned}$$

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

The vector $\text{ret}'_{it} = (stI2_{it}, stII3_{it}, ftII3_{it})$ include three mutually exclusive dummies to capture all the possible combinations of time and repetitions. $stI2$ takes value 1 if

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

the student failed at $t = 1$ and has to repeat level I at $t = 2$. *stII3* is 1 if the student has to repeat level II at $t = 3$. *ftII3* is 1 for a student who repeated the first level at $t = 2$ and can undertake second level for the first time at $t = 3$. The specification in (1.17) allows cognitive skills to have different effects on the flow utility of students with differing retention history. This specification also allows school, peers, and individual characteristics to have different effects depending on the level of cognitive skills, while remaining parsimonious on the number of parameters to estimate.[21]

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

$$U_{it}^{A} = \phi_A \mathrm{E}_{i,t}(C_{\mathrm{II}}) + \mathrm{ret}'_{A,it}\theta_{A,r} + x'_i\theta_{A,x} + p'_{it}\theta_{A,p} + s'_{it}\theta_{A,s} + \tag{1.19}$$

$$+ \mathrm{E}_{i,t}(C_{\mathrm{II}})\Big[\mathrm{ret}'_{A,it}\kappa_{A,r} + \kappa_{A,x}(x'_i\theta_{A,x}) + \kappa_{A,p}(p'_i\theta_{A,p}) + \kappa_{A,s}(s'_{it}\theta_{A,s})\Big] + \varepsilon_{i,t} =$$

$$= (\phi_A + \widetilde{q}'_{it}\kappa_A)\mathrm{E}_{i,t}(C_{\mathrm{II}}) + q'_{it}\theta_A + \varepsilon_{i,t}. \tag{1.20}$$

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

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

1.3.7 Choices and optimization

In each period students make a schooling decision taking in account their flow utility and expected future utility; individuals are assumed to be forward-looking and choose the sequence of actions which yield the highest expected value.
The one-period discount factor is δ. I use u_{it} to denote the utility at time t. Recall that students know all present and futures covariates while signals and shocks to preferences are random variables.

At the end of lower secondary education. For those who graduated, the utility of pursuing further education is simply the flow utility in (1.19), with $t = 2$ if retention never took place or $t = 3$ if the student was retained in either first or second level. Therefore:

$$u_{it}|(\mathrm{grad}_{i,t} = 1) = \max\Big\{0\,,\; U_{it}^{A}\Big\} = \max\Big\{0\,,\; v_{it}^{A} + \varepsilon_{it}\Big\}, \tag{1.21}$$

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

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

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

$$u_{i2}|(\text{grad}_{i2} = 0) = \max\left\{0\,,\, U_{i2}^M + \delta\,\Pr(\text{grad}_{i3} = 1)\text{E}_{i,2}(u_{i3}|\text{grad}_{i3} = 1)\right\} = \\ = \max\left\{0\,,\, v_{i2}^M + \varepsilon_{i2}\right\}. \tag{1.22}$$

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

$$u_{i1}|(\text{failI}_i = 0) = \max\Big\{0\,,\, U_{i1}^M|(\text{failI}_i = 0) + \tag{1.23} \\ + \delta\Big(\Pr(\text{grad}_{i2} = 1)\text{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 1)) + (1 - \Pr(\text{grad}_{i2} = 1))\text{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 0)\Big)\Big\} \\ = \max\left\{0\,,\, v_{i1}^M|(\text{failI}_i = 0) + \varepsilon_{i1}\right\},$$

$$u_{i1}|(\text{failI}_i = 1) = \max\left\{0\,,\, U_{i1}^M|(\text{failI}_i = 1) + \delta\text{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 0)\right\} = \\ = \max\left\{0\,,\, v_{i1}^M|(\text{failI}_i = 1) + \varepsilon_{i1}\right\}. \tag{1.24}$$

1.3.8 Identification

As common for this type of dynamic discrete choice models (e.g., see Rust (1987) and Arcidiacono et al. (2016)), identification of the flow utility parameters relies on the distributional assumptions imposed on the idiosyncratic shocks, the normalization of the outside option, and the discount factor δ, which is set equal to 0.95 throughout the paper.[22]

Under the assumptions on the parameters already discussed in Subsection 3.4.2, the identification of the grade equations relies on the assumption that educational choices only depend on students ability through their belief. In fact, evaluations at time $t \geq 2$ are only observed for individuals who chose to continue their education; to the extent that the choice depends on their ability, this raises a selection issue. The next paragraphs provide the intuition of why parameters of the grade equations can be consistently estimated.

Consider the following regression for external evaluations in level II and $t \in \{2, 3\}$,

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

based on equation (1.11) in Subsection 3.4.2. Let assume for now that posterior belief $\mathrm{E}_i(h)$ has been computed for each student.

$$r_{\mathrm{II},it} = o_{\mathrm{II}} + z'_{it}\beta_{\mathrm{II}} + z'_{\mathrm{I},i}\widetilde{\kappa}_{\mathrm{II}} + \lambda_{\mathrm{II}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{r_{\mathrm{II}},it}, \tag{1.25}$$

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

Similarly, ordinary least square would consistently estimate the reduced form parameters of the following regression (based on equation 1.12):

$$g_{\mathrm{II},it} = \nu_{\mathrm{II}} + z'_{it}(\mu\beta_{\mathrm{II}} + \gamma) + z'_{\mathrm{I},i}\mu\widetilde{\kappa}_{\mathrm{II}} + \mu_{\mathrm{II}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{g_{\mathrm{II}},it}, \tag{1.26}$$

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

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

$$g_{\mathrm{I},it} = \nu_{\mathrm{I}} + z'_{it}(\beta_{\mathrm{I}} + \gamma) + z'_{0,i}\widetilde{\kappa}_{\mathrm{I}} + \mu_{\mathrm{I}}\mathrm{E}_i(h) + \widetilde{\epsilon}_{g_{\mathrm{I}},it}, \tag{1.27}$$

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

Finally, ordinary least square applied to

$$r_{0,i} = z'_{i0}\beta_0 + h_i + \widetilde{\epsilon}_{r_0,i} \tag{1.28}$$

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

So far the identification of the parameters rested on the simplifying assumption that belief $\mathrm{E}_i(h)$ have been already computed. In fact, to perform the bayesian updating one should know the variance σ of the ability h and the variances of the errors in the grade equations $(\rho_0^r, \rho_{\mathrm{I}}^g, \rho_{\mathrm{II}}^g, \rho_{\mathrm{II}}^r)$. Those are identified from the history of signals, particularly the covarariance of evaluations within and overtime. In particular, σ is inferred from the covariance of the residuals of $g_{\mathrm{II},it}$ and $r_{\mathrm{II},it}$ on the observable regressors. The variance of each type of residuals is a linear function of σ and of the variance of the relevant error, thus the latter can be retrieved after estimating σ.[23]

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

1.4 Estimation

This section derives the likelihood of the model described in Section 2.2 and discusses its estimation.

1.4.1 Total individual likelihood

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

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

$$L_i = L(d_i, \text{grad}_i, o_i) = L(d_{i1}, ..., d_{iT_d}, \text{grad}_{i1}, ..., \text{grad}_{iT_{\text{gr}}}, g_{i1}, ..., g_{iT_d}, r_{i1}, ..., r_{iT_r}) \tag{1.29}$$

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

$$\begin{aligned} L_i = \int L(r_{i,0}|h)L(\text{grad}_{i,1}|h, r_{i,0})L(g_{i,1}|h, r_{i,0})(L(d_{i,1}|h, r_{i,0}, g_{i,1})...L(d_{iT_d}|h, o_i, d_{i,1}, ..., d_{it_d-1})\, \phi(h)dh = \\ \Big(L(d_{i,1}|r_{i,0}, g_{i,1})...L(d_{i,T_d}|o_i, d_{i,1}, ..., d_{i,T_d-1})\Big) \times \Big(L(\text{grad}_{i,1}|r_{i,0})...L(\text{grad}_{i,T_{\text{gr}}}|o_i, d_{i,1}, ..., d_{i,T_d-1})\Big)\times \\ \times \int L(o_{i,T_d}|h, d_{i,1}, ..., r_{i,0}, ...)...L(r_{i,0}|h)\, \phi(h)dh \end{aligned} \tag{1.30}$$

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

$$\log L_i = \log L_{i,d} + \log L_{i,\text{grad}} + \log L_{i,o} \tag{1.31}$$

Once coefficients of $\log L_{i,o}$ have been estimated, they can be used in the other components. In particular beliefs on cognitive skills can be computed for each student at any point in time and used as regressors. Then one has to estimate straightforward logit models for the probabilities, and a model of dynamic choices with logistic errors. More details are provided in subsections 1.4.3 and 1.4.4.

On the other hand maximizing the likelihood $\log L_{i,o}$ would be computationally costly because of the integration of h. Following James (2011) and Arcidiacono et al. (2016) I use an Expectation-Maximization (EM) algorithm to overcome this issue. I summarize the implemented approach in Subsection 1.4.2.

1.4.2 Cognitive skills

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

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

$$\arg\max_{\zeta,\sigma} \sum_i \log L(o_i; \zeta, \sigma) = \arg\max_{\zeta,\sigma} \sum_i \log \int L(o_i; \zeta, \sigma|h)\boldsymbol{\phi}(h)\, dh \tag{1.32}$$

The main point behind this application of the EM algorithm is that if $\widehat{\zeta}$ is a maximizer for (3.1), then it also solves

$$\arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma|h)\boldsymbol{\psi}_i(h)\, dh \tag{1.33}$$

Therefore for a given value of σ, $\widehat{\zeta}$ can be retrieved using (3.8) rather than (3.1). $\widehat{(\zeta, \sigma)}$ 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 (3.8).

Appendix (1.C) provides a more detailed theoretical motivation. Next paragraphs describe the estimation procedure.

E-step

At step k, posterior distribution $\boldsymbol{\psi}_i^k(h)$ are computed for every students using all the observed evaluations, and the parameters $(\zeta^{k-1}, \sigma^{k-1})$ estimated in the previous iteration. Let $\mathrm{E}_i^k(h)$ be the individual posterior belief for h at iteration k and $\omega_i^k(h)$ the posterior variance. Moreover, at the end of E-step, we also update the estimate for the population variance; this new σ^k will be used at the beginning of next step $k+1$. The updating formula for σ^k is retrieved using the law of total variance:[24]

$$\sigma = \mathrm{E}\left(\omega_i(h) + \mathrm{E}_i(h)\mathrm{E}_i(h)'\right), \tag{1.34}$$

The sample equivalent at step k is computed as

$$\widehat{\sigma}^k = \frac{1}{N}\left(\omega_i^k(h) + \mathrm{E}_i^k(h)^2\right) \tag{1.35}$$

M-step

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

$$\arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh \tag{1.36}$$

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

$$\sum_i \int \log L(o_i; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh = \tag{1.37}$$

$$= \sum_i \left(\sum_t \int \log L(r_{it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \sum_t \int \log L(g_{it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh \right) = \tag{1.38}$$

$$= \sum_i \int \log L(r_{0,i}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \sum_{it} \int \log L(g_{I,it}; \zeta, \sigma^{k-1}|h)\boldsymbol{\psi}_i^k(h)\, dh + \tag{1.39}$$

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

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,

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

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

the above expression can be derived as follow:[25]

$$
\begin{aligned}
-\sum_i logL_i \; &= \sum_i \frac{1}{2}\log(2\pi\rho_{r,0}) + \frac{1}{2\rho_{r,0}}\left(\omega_i^k + \left(r_{0,i} - (\mathrm{E}_i^k(h) + z'_{i0}\beta_0)\right)^2\right) + \\
&+ \sum_{it} \frac{1}{2}\log(2\pi\rho_{g,\mathrm{I}}) + \frac{1}{2\rho_{g,\mathrm{I}}}\left(\omega_i^k + \left(g_{\mathrm{I},it} - \left(\nu_1 + \mu_{\mathrm{I}}\mathrm{E}_i^k(h) + z'_{it}(\beta_{\mathrm{I}} + \gamma) + \kappa_{\mathrm{I}} I_{0,i}\right)\right)^2\right) \\
&+ \sum_{it} \frac{1}{2}\log(2\pi\rho_{r,\mathrm{II}}) + \frac{1}{2\rho_{r,\mathrm{II}}}\left(\omega_i^k + \left(r_{\mathrm{II},it} - \left(o_{\mathrm{II}} + \lambda_{\mathrm{II}}\mathrm{E}_i^k(h) + z'_{it}\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i}\right)\right)^2\right) \\
&+ \sum_{it} \frac{1}{2}\log(2\pi\rho_{g,\mathrm{II}}) + \frac{1}{2\rho_{g,\mathrm{II}}}\left(\omega_i^k + \left(g_{\mathrm{II},it} - \left(\nu_{\mathrm{II}} + \mu\lambda_{\mathrm{II}}\mathrm{E}_i^k(h) + \mu(z'_{it}\beta_{\mathrm{II}} + \kappa_{\mathrm{II}} I_{\mathrm{I},i}) + z'_{it}\gamma\right)\right)^2\right)
\end{aligned} \tag{1.41}
$$

The total likelihood in (1.41) is the sum of four parts, one for each type of evaluations. Some students may contribute twice to the likelihood of an evaluation if they are retained, or they may not have some of them (if they dropout or do not take the final external evaluation).

If all the regressors were time invariant (i.e. if $I_{\tau,i}$ were not in the equations), the joint estimation of (1.41) would be completely equivalent to separately estimate the coefficients for r_0, then for g_{I}, and finally jointly estimates coefficients of r_{II} and g_{II}. Conversely the presence of time varying regressors makes all the four parts interdependent because past regressors have an indirect effect on evaluations in the following periods. Therefore a joint estimation is the most efficient. In practice, I found the following two-step MLE to be a good compromise between efficiency and speediness of the computations:

1. *Parameters for external evaluation at the end of primary school.*
    - Perform OLS regressions of $r_{0,i} - h_i$ over $z_{i,0}$. This provides us with updated estimates β_0^k, and allows the computation of $I_{i,0}^k = s_{i,0}\beta_{s,0}^k$, which is used in next step.

[25] It is easy to see how to derive (1.41) from (1.39). For instance the contribution of the first nation-wide test is given by:

$$
\begin{aligned}
&\int \log L(r_{0,i};\zeta,\sigma^{k-1}|\eta)\psi_i^k(h)\,dh = \int \log\left(\frac{1}{\sqrt{2\pi\rho_0^r}}\exp\left(-\frac{(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2}{2\rho_0^r}\right)\right)\psi_i^k(h)\,dh = \\
&= \int\left(-\frac{1}{2}\log(2\pi\rho_0^r) - \frac{1}{2\rho_0^r}(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2\right)\psi_i^k(h)\,dh = \\
&= -\frac{1}{2}\log(2\pi\rho_0^r) - \frac{1}{2\rho_0^r}\mathrm{E}_i^k\left((r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2\right) = \\
&= -\frac{1}{2}\log(2\pi\rho_0^r) - \frac{1}{2\rho_0^r}\left(\mathrm{Var}_i^k\left(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0}\right) + \left(\mathrm{E}_i^k(r_{0,i} - h - x'_i\beta_{x,0} - p'_i\beta_{p,0})\right)^2\right) = \\
&= -\frac{1}{2}\log(2\pi\rho_0^r) - \frac{1}{2\rho_0^r}\left(\omega_i^k + \left(r_{0,i} - \mathrm{E}_i^k(h) - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0}\right)^2\right)
\end{aligned} \tag{1.40}
$$

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

$$\mathrm{Var}(\epsilon_{r_0,i}) = \mathrm{E}(\epsilon^2_{r_0,i}) = \mathrm{E}(\mathrm{E}((r_{0,i} - h_i - z'_{i0}\beta_0|r_{0,i})^2)) = \tag{1.42}$$

$$= \mathrm{E}\left(\int (r_{0,i} - h - z'_{i0}\beta_0)^2\, \psi_i(h)\, dh\right) = \tag{1.43}$$

$$= \mathrm{E}\left[\mathrm{Var}_i(h) + \left((r_{0,i} - \mathrm{E}_i(h) - z'_{i0}\beta_0)^2\right)\right] \tag{1.44}$$

which can be estimated from the sample as

$$\rho^k_{r,0} = \frac{\sum_i \left(\omega^k_i + (r_{0,i} - E^k_i(h) - z'_{i0}\beta^k_0)^2\right)}{N} \tag{1.45}$$

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

1.4.3 Retention and graduation probabilities

Using the estimated parameters $\widehat{\zeta}$ and individual posterior distributions of ability $\widehat{\psi}_i$ at each point in time, we can compute beliefs on cognitive skills and use them to estimate retention and graduation probabilities, following the approach described in Section 1.3.5. I estimate two separate logit equations, one for retention in first period and one for graduation/retention in second period. The individual probability of graduation enters the student's maximization problem, while the probability of retention in first period does not, given that it happens before any decision has to be taken. However the estimation of the latter is necessary for simulations and conterfactual analyses.

1.4.4 Dynamic choices

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

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

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

$$\mathrm{E}_\varepsilon(u_{it}|v^A_{it}) = \log\big(\exp(v^A_{it}) + 1\big) = \log\left(\exp(\phi_A \mathrm{E}_{i,t}(C_{\mathrm{II}}) + q'_{it}\theta_A + Hq'_{it}\kappa_A) + 1\right) \tag{1.46}$$

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:

$$\mathrm{E}_{i,t}(h) = (\widehat{h}_{it-1}, \mathbf{s}'_t)\mathbf{c}_t = c_{0,t}\widehat{h}_{it-1} + c_{1,t}s_{r,t} + c_{2,t}s_{g,t} \tag{1.47}$$

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

$$E_{i,t-1}(u_{it}|\mathrm{grad}_{it} = 1) = \int \log\left(\exp(v_{i,A}) + 1\right) \cdot \widehat{\psi}_{it-1}(s_{g,t}, s_{r,t})\, d\mathbf{s}_t = \tag{1.48}$$

$$= \int \log\Big[\exp\left((\phi_A + \widetilde{q}'_{it}\kappa_A)\left(k_{\mathrm{II}}I_{t-1} + z'_{it}\beta_{\mathrm{II}}\right) + q'_{it}\theta_A\right) \cdot \exp\Big((\phi_A + \widetilde{q}'_{it}\kappa_A)(c_{0,t}\widehat{h}_{it-1})\Big) \cdot$$
$$\cdot \exp\left((\phi_A + \widetilde{q}'_{it}\kappa_A)(c_{1,t}s_{r,t} + c_{2,t}s_{g,t})\right) + 1\Big] \cdot \widehat{\psi}_{it-1}(s_{g,t}, s_{r,t})\, d\mathbf{s}_t$$

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}:

$$v^M_{i2}(\widehat{h}_{i2}, z_{i2}) + \varepsilon_{i2} = U^M_{i2} + \delta\widehat{\Pr}(\mathrm{grad}_{i3} = 1)\int \log\big(\exp(v^A_{i3}) + 1\big) \cdot \widehat{\psi}_{i2}(s_{g,3}, s_{r,3})\, d\mathbf{s}_3 \tag{1.49}$$

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:

$$v^M_{i,1}|(\text{failI}_i = 1) \; + \varepsilon_{i,1} = U^M_{i,1} + \delta E_{i,1}(u_{i2}|\text{grad}_{i2} = 0) \tag{1.50}$$

$$\begin{aligned} v^M_{i,1}|(\text{failI}_i = 0) \; + \varepsilon_{i,1} = U^M_{i,1} + \delta\Big(& \Pr(\text{grad}_{i2} = 1)\mathrm{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 1)+ \\ & + (1 - \Pr(\text{grad}_{i2} = 1))\mathrm{E}_{i,1}(u_{i2}|\text{grad}_{i2} = 0)\Big) \end{aligned} \tag{1.51}$$

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:

$$p_{it}(d_{it} = 1|d_{it-1} = 1) = \frac{\exp(v_{it})}{1 + \exp(v_{it})} \tag{1.52}$$

I maximize the total loglikelihood to estimate the parameters, following Rust (1987).[26]

1.4.5 Missing external evaluations

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

1.4.6 Standard errors

Standard errors are estimated using a bootstrap procedure with 100 replications. Let N_s be the number of students in the sample enrolled in school s. In each replication, for each school s, N_s individuals are sampled with replacements.

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

1.5 Results

In this section I present the results of the estimation. In subsections 1.5.1 to 1.5.3 I discuss the estimated parameters; in Subsection 1.5.4 I discuss the fit of the model. The main findings are the following:

- *Cognitive skills.* About half of the total variance of cognitive skills is due to unobserved ability, however evaluations are informative and posterior individual variance shrink rapidly.[27] Parental education is the most important observed determinant of cognitive skills accumulation: having both parents with tertiary degree rather than primary education is associated with a 1 s.d. deviation improvement. School environment is quite relevant as well: being with higher ability peers increases evaluations (more than 0.1 s.d. for peers 1 s.d. above average), and there is large variation in the school effects (e.g. the difference between the school effect at the 75 percentile and the one at the 25 percentile is 0.35 s.d.). School effects are quite homogeneous across individual characteristics.

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

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

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

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

1.5.1 Cognitive skills and evaluations

As reported in the first entry of Table 1.3 (Panel A), the estimated variance of the individual unknown ability is 0.277. As explained in Section 3.4.2, the total variance of cognitive skills in a given level does not have any direct interpretation; however, given the assumption that h is uncorrelated with the regressors, it is possible to decompose it in variance due to observable characteristics, and variance due to unknown ability. The remaining entries of Table 1.3 (Panel A) display the share of total variance in each level due to ability. In all levels about half of the variance in cognitive skills is due to the variance in ability; more specifically, about 49% before starting middle school and in the first level, and about 58% in the second level.

Panel B contains posterior variance in each period for every possible set of signals. Evaluations are quite informative, in fact the posterior variance at time $t = 1$ is just 0.06; at $t = 2$ it shrinks to 0.036 for retained students and 0.029 for students who are promoted and receive both internal and external evaluations; it is further reduced to about 0.02 for retained students who stay in middle school up to $t = 3$. Posterior variance is just slightly larger for students who did not undertake external evaluations in level II.

Tables 1.4 and 1.5 present estimates for the parameters governing internal and external evaluations, in periods 0 and I, and in period II, respectively. For the convenience of the reader, the tables also report separately the contributions of β and γ to internal evaluations in level I and II. Since the evaluations have been standardized, the coefficients of the individual characteristics can be interpreted as standard deviation changes of the relevant evaluation. Being a female is associated with a large premium for internal evaluations (0.33 in first level and 0.48 in second level), while boys and girls perform similarly in external evaluations (and thus there aren't relevant differences in their cognitive skills accumulation everything else equal). External evaluations are about 0.16 s.d. lower for non-Spanish students; the gap is small for internal evaluations. In fact, being younger when entering primary education is a disadvantage, although the gap is decreasing over time. Being born at the beginning of January rather than at the end of December is associated with an increase of about 0.25 s.d. in the test score before starting middle school and with a increase of about 0.13 s.d. at the end of it. Being retained in primary school, i.e. starting middle school with a one-year delay is associated with a large disadvantage in performances (up to -0.66 in external evaluations at the end of middle school).

Results show that parental background is a fundamental determinant of cognitive skills. Having better educated parents is associated with large improvements in performances, with quite similar effect for internal and external evaluations; for instance, having a mother with tertiary rather than primary education increases performances of about 0.6 s.d., having a highly educated father increases performances of about 0.3 s.d. This means that a student with highly educated parents is expected to receive evaluations almost 1 s.d. higher than a classmate with identical characteristics and unobserved ability, but whose parents only have primary education.

Repeating a level for a second time has a positive effect on cognitive skills, especially in second level. In fact final external evaluations are more than 0.3 s.d. larger for retained students everything else equal (the effect is 0.17 in first level).

The empirical analyses use a polynomial of degree 3 with orthonormal components for each peer regressor. To facilitate the interpretation of the results, the tables show the effect on evaluations associated with having peers 1 s.d. above the mean value rather than at the mean. Estimated coefficients are reported in Table 1.15. Peers' average evaluations at the end of primary school, an *ex ante* measure of average peer quality, is the peers' characteristic which affects cognitive skills development the most. 1 s.d. increase of this measure is associated with almost 0.1 s.d. increase in skills at the end of middle school (0.18 s.d. increase in cognitive skills in first period). On the other hand, the positive effect is completely offset in internal evaluations. This is aligned with the finding in Calsamiglia and Loviglio (2018) that teachers' evaluations are deflated in the presence of better quality peers, i.e. that teachers are somewhat "grading on a curve". Average parental education among peers has a positive effect on cognitive skills in the first level (about 0.1 s.d. deviation), while the effect is positive but quite small in second level. The other peer characteristics appear to have limited effects.

To summarize the estimated school effects, Tables 1.4 and 1.5 display the difference between the school dummy coefficients at the 75th percentile and at the 25th percentile, and the difference between school dummy coefficients at the 80th and 20th percentile. There is a sizable variation across schools in their effects; the interquantile range for the effect on evaluations through cognitive skills is more than 0.3 s.d. in the first level and almost 0.3 in second level. The interquantile range for the inflation γ_S of internal evaluations is sizeable as well (0.4). However the school inflation effect $\mathcal{I}$ and the school effects on cognitive skills ($\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$) have large negative correlation (about -0.7). In other words, schools which improve the cognitive skills the most tend to have stricter grading policy.[28] Therefore the correlation of the ranking of schools based on school effect

[28] This is coherent with the the evidence in Calsamiglia and Loviglio (2018) that schools in which

on external evaluations and total school effect on internal evaluations is low, although positive.

The estimation includes interaction between the vector of school effects and the dummies for gender, nationality, and parental education. The estimated parameters are reported in Table 1.15, while Tables 1.4 and 1.5 report for each individual characteristic the change due to an increase of 1 s.d. in the school effect. Overall the estimated interaction effects are quite small: schools appear to affect similarly cognitive skills development of students with differing gender, nationality, and parental background.

1.5.2 Retention and graduation

Estimates of the parameters governing the two logit models for retention in first level and retention/graduation in second level are reported in Table 1.6.[29] Not surprisingly, higher cognitive skills decreases the probability of retention in both periods. Being female and having better educated parents decrease the probability of retention in both periods above and beyond the cognitive skills level.[30]

There is sizable variation in school effects on retention, and they are correlated across levels, i.e. in schools that have a stricter retention policy in the first level it is also more difficult to graduate in the second level. For retention in the first level, the effect of being in a school at the 25 percentile rather than at the 75 percentile is the same as increasing cognitive skills by 0.34 s.d.. For graduation, being in a school at the 75 percentile rather than at the 25 percentile has the same effect of increasing skills of 0.5 s.d.[31]

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

1.5.3 Flow utilities

Table 1.7 presents the estimates of the flow utility parameters. Beliefs about cognitive skills have a large positive effect on both the choice of continuing middle school and the choice of enrolling in high school.

Peers' average incoming test score has a negative effect on the choice of staying in middle school (beyond their positive contribution to cognitive skills development). This

average external evaluations are higher exhibit stricter grading policy.

[29]The table has the same structure of Table 1.4: for peer regressors it shows the effect on the average student in the sample of increasing the variable of one standard deviation; for school effects it reports differences between percentiles of the distribution.

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

[31]These figures are computed using the total variance of cognitive skills reported in Table 1.3.

may be related to ranking concerns: for a given level of cognitive skills, students at the margin of dropout may have a further reason to leave school if they dislike being among the worst in the class.[32]

School effects exhibit sizable differences both for the choice of staying in middle school and the choice of enrolling in high school. For the choice of staying in middle school, being in a school at the 75 percentile of school effects rather than at the 25 percentile impacts the flow utility as much as improving cognitive skills by 0.45 s.d. Given that the coefficient of the interaction between cognitive skills and school effects is almost 0, this difference is about constant for all level of cognitive skills. For the choice of enrolling in high school, the difference between being in a school at the 75 percentile rather than at the 25 percentile is equivalent to an increase of cognitive skills by 0.67 s.d. for someone with average level of cognitive skills. However the interaction coefficient is negative, therefore the gap is wider for children with below average cognitive skills and smaller for above average students. Figure 1.1 helps clarifying this point. The figure plots the school effects at each level of cognitive skills: schools with a higher coefficient for the dummy have a larger intercept and a smaller slope. The point in which the lines intersect is about 2 standard deviation above the average value of cognitive skills at the end of second level: students with such a high level of cognitive skills are extremely likely to enroll in high school anyway, therefore it is sensible to focus on the part of the graph which lies to the left of the intersection. There, at any given level of cognitive skills, students have a lower payoff from the choice of enrolling in high school if they attended a school with low effect. This gap increases when cognitive skills decreases. In other words, schools affect the flow utility of students in the lower tail of the distribution of cognitive skills the most.

All dummies for the retention events have negative coefficients. Interestingly, their interactions with cognitive skills also have negative coefficients, thus the total coefficient of cognitive skills for retained students is smaller than for students who are progressing regularly. In other words, being retained decreases the weight that individuals give to cognitive skills when making their choices. Figure 1.2 illustrates how the retention status affects the effect of cognitive skills on flow utility of enrollment in high school. At any level of skills retained students have a lower payoff, but the gap is is wider at higher levels of cognitive skills: an increase in skills improves less the utility of retained than of non-retained students.[33] Being retained in first or in second level have quite similar effects.

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

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

As discussed in previous Subsection 1.5.2, retention probability is decreasing in cognitive skills, however many students with cognitive skills below average, but not particularly low, face a sizable probability of being retained at some points, especially if they are male; retention may particularly discourage those guys from pursuing further education.

1.5.4 Fit of the model

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

Table 1.8 reports empirical frequencies ("data" columns) and frequencies predicted using the model ("model" columns) for the following events: choice of staying in school at time 1, graduation, enrollment in high school, retention in first level, retention in second level. Frequencies are computed over the entire sample at time 1. Table 1.16 in the appendix reports similar statistics computed only on the subsample of the initial population who reached the relevant stage for the event to take place (e.g. enrollment on high school on the subgroup of students who completed middle school). The first line of each table contains frequencies for the overall sample, while the following rows contain the same type of information by relevant subgroups (e.g. parental background, quality of peers...). The predicted choice of staying in school, graduation rate, and choice of enrolling in high school are very close to the empirical one, both in the overall sample and by subgroups. The retention rate in first level is also similar to the empirical one, while the retention rate in second level is somehow higher (about 1.7 p.p. more).

Next, I investigated how evaluations and events simulated by the model replicate the patterns observed in the data. For instance, Figure 1.3 plots the share of students who chose to stay in school at $t = 1$ by quantile of their test score at $t = 0$; Figure 1.4 plots their enrollment in high school, again by quantile of the initial test score. The model predictions mimic the empirical outcomes quite well. Other evaluations and choices exhibit similar patterns.

like high school less in absolute terms, but that they value it relatively less compared with the outside options.

1.6 School environment and parental background

In this section I study how the school environment affects outcomes of children with different parental background. In particular I focus on students with low educated parents (both attained at most lower secondary education) and on students with highly educated parents (both have tertiary education). As discussed in Section 1.5.1, having more educated parents is associated with a larger growth of cognitive skills, even if the level of unobserved ability is the same. Moreover, Table 1.9 shows that on average students with low and high parental background experience a different school environment.

In particular, panel A shows that students with low educated parents have peers with less educated parents, who are more likely to be non-natives, and to have experienced retention in the past. Looking at the distribution of mean peers' evaluation at the end of primary school, the expected value for a student with low educated parents is at the 23 percentile, while it is at the 64 percentile for children of highly educated parents. Similarly, the share of immigrant classmates is at the 73 percentile rather than at the 45.

Panel B displays average school effects for the two parental backgrounds; the typical school for a student with low educated parents has a somewhat more generous grading policy, a smaller propensity to retain in first level, larger direct effect on the choice of staying in school, and contribute slightly less to cognitive skills accumulation in first level. Average values of school effects for second level are close enough for the two types. Figure 1.5 confirms that there is some positive correlation among school effect on choice of staying and share of students with low educated parents enrolled in the school, while the share of students with low educated parents is uncorrelated with school effect on choice of high school, and has small negative correlation with school effect on cognitive skills. Overall differences in average school effects appear less relevant than those for peers, possibly with the exception of school effect on dropout. Importantly, almost all schools have both students with low educated and high educated parents.[34]

Results in Subsection 1.5.1 show that both peer quality and school effects on top of observable peer characteristics matter for cognitive skills accumulation. Moreover they show that school effect is almost the same for students with high and low parental background. Therefore effects due to school environment are about the same for all students in a given class. If on average school environment affect differently students with low or highly educated parents, this is due to their different allocation across schools.

On the other hand, subsections 1.5.2 and 1.5.3 show that schools affect students' attainments and choices beyond their contribution to cognitive skills. School environments

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

that are not particularly effective in boosting cognitive skills may have a positive effect on choices of pursuing further studies, through at least two channels. First, school effects on cognitive skills and choices are negatively correlated. Second, for a given level of cognitive skills students are more likely to dropout if they have higher performing peers.

Importantly, school environment may affect differently classmates who differ in their individual characteristics or ability. This is due to the non linearity of the probability functions: the effect on the probability of a change in one of the regressor depends on the initial level of the underlying utility.[35] Moreover, the flow utilities as well are non linear, because they include interactions between school environment and cognitive skills, and between retention history and cognitive skills (and schools differ in their propensity to retain students). Therefore, the school environment may matter differently for educational choices of students with low or highly educated parents even if they are attending the same class.

I use the model and the estimated parameters to study performance, choices, and attainments of students with low and high parental background in their typical school environment or in counterfactual environments (Subsection 1.6.1). Then, I simulate outcomes of a given student in each school of the sample to quantify the variance in performances and attainments due to the school environment (Subsection 1.6.2). Finally, I explore differences in educational patterns of retained and non-retained students who have identical cognitive skills before retention (Subsection 1.6.3).

More specifically, for a given set of individual and school characteristics, I create 10000 fictitious students, and I draw shocks to evaluations, preferences, and retention events. I can then compute the educational pattern of each agent and estimate expected outcomes of an individual with a given set of individual characteristics and school environment. To make results fully comparable, I keep constant all individual characteristics beside parental education. In particular, the reference characteristics for the results discussed in this paper are: male, Spanish, born in the middle of the year, began middle school at 12

[35] As explained in Section 1.4.4, given the assumptions that shocks to preferences follow a logistic distribution, the probability of undertaking choice d_t at time t is $\Lambda(v_t(z)) = \frac{\exp(v_t(z))}{1+\exp(v_t(z))}$, where z is the vector of individual and school variables and v is the expected utility (i.e. the flow utility at time t before observing the shock to preferences plus the expected utility from the future). The marginal effect of a change in one of the regressor z_k is

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

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

years old; the unobserved ability is set to its average value of 0.[36] For brevity, I will call *type L* the students with such characteristics and low educated parents, and *type H* the students with such characteristics and highly educated parents.

1.6.1 Peer and school effects

Table 1.10 summarizes outcomes of type L and type H when they are in their typical school environment, and under various counterfactual scenarios in which elements of the school environment of type H are assigned to type L and vice-versa. In the baseline simulation (columns (1)) each type attends a school with the average peers and school characteristics of his type. The first column of the table reports also results for a student with average parental background for comparison. Not surprisingly, type L receives below average internal and external evaluations throughout his studies, while type H's performances are above average. For instance, they enter middle school with external evaluations of -0.35 and 0.50 respectively, and those who graduate score on average -0.40 and 0.70 respectively. Type L students have 59% chances of being retained in first period, and almost 15% of dropping out immediately; more than 30% of them do not complete middle school and only 25% overall enroll in high school (about 36% of those who complete middle school). Conversely, only 4% of type H students is retained in first period, 0.7% dropout immediately, and only 1% do not graduate; about 94% of students enroll in high school (95% of those who graduate).

The remaining columns of Table 1.10 show results of conterfactual simulation in which a given type is enrolled in a school with typical school effects of the other type (columns (2)), or peers of the other type (columns (3)), or both (columns (4)). The comparison of the outcomes in the counterfactual and in the baseline specification allows us to gain a deeper understanding of the role plaid by the school environment on students' outcomes.

When type L attends a school with the average school effects of type H, he acquires a slightly larger level of cognitive skills – the final evaluation is about 0.05 s.d. higher. However he is slightly more likely to be retained, and he choose to dropout more both at time $t = 1$ and at time $t = 2$. Therefore he is about 5 p.p. less likely to graduate; probability of enrolling in high school conditional on graduation increases by 1.5 p.p. Type H would have a small drop in cognitive skills if he attended the typical school of the low type, but his attainment would be virtually unaffected. Changes in the grade inflation coefficient $\mathcal{I}$ and in the cognitive skills coefficients $\mathcal{A}_{\text{I}}$ and $\mathcal{A}_{\text{II}}$ largely compensate

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

each other, therefore for both types internal evaluations are almost unchanged.

A change in peers composition has more dramatic effects, as shown in columns (3) of the table. For type L, the improvement in parental background and test score of peers implies an increase of cognitive skills of about 0.3 s.d. at time $t = 1$. A slightly lower but still sizable improvement takes place also at the end of middle school. Conversely internal evaluations are similar in first level and somewhat lower in second level because the negative effect due to "grading on the curve" more than compensate the improvement in cognitive skills. New peers decrease the probability of retention, which is reduced by half. The boost in human capital more than compensate the negative effects that having better performing peers has on the choice of remaining in education; in fact dropout rate plummets from almost 15% to about 7% at time $t = 1$ and from 23% to 15% at time $t = 2$, and the graduation rate raises from about 69% to about 84%. Moreover enrollment in high school conditional on graduation increases from 36% to 52%.

Type H experiences a symmetric drop in performances, and an increase of retention probability at $t = 1$ from 4% to 13%. On the other hand, dopout rate at $t = 1$ changes from 0.7% to 2.2%, graduation rate from 99% to 97%, enrollment in high school from 94% to 87%: the underlying flow utilities for this type are so high that variations in them lead to relatively small changes in probabilities.

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

Summarizing, type L would benefit from attending the typical school in which type H is enrolled mainly because "better peers" would increase his cognitive skills, and this would have a large positive impact on his choices and attainments. The improvement in cognitive skills would not raise internal evaluations due to the counterbalancing effect of "grading on the curve". Conversely, the average school environment beyond peers is not very different for the two types, and it would not dramatically change type L outcomes; if anything, on average the school which he is already attending is better suited to increase his motivation to acquire further education on top of its level of cognitive skills. However, results in Section 2.4.1 show that there is large variance across schools in their school effects beyond peers, therefore the same type of student may experience a quite different educational path depending upon the specific institution in which he is enrolled. I will study the relevance of school environment school by school in next Subsection 1.6.2.

Results in this section and parameters estimated in Section 2.4.1 suggest that the event of retention can be important for subsequent choices and attainments; Subsection 1.6.3 will provide further insights.

1.6.2 Variation of outcomes across schools

To quantify the importance of the school environment on attainments and choice, I use the model to predict for each school the outcomes that a given type of student would have if enrolled there. Table 1.11 presents results for type L and type H. As usual, for each type I created 10000 fictitious students, and I drew shocks to evaluations, preferences, and retention events; then I simulate their outcomes in each of the 44 schools in the sample. I use the average peer characteristics in the school for the peer regressors; results are therefore representative of the outcomes that a student of a given type would have enrolling in one of the 44 schools in the sample. For comparison, I also focus on peer and school effects separately: I perform a simulation in which peer characteristics are set at their average values in the sample (Table 1.20) and a simulation in which school effects are set at their average in the sample (Table 1.21). Note that in all these simulations type L and type H are exposed to exactly the same environment in a given school: this allows me to quantify how relevant is the school environment for each of the type.

The left half of Table 1.11 contains results for the student with low parental background, the right half contains results for the student with highly educated parents. The first column of each part contains median outcomes across school. Patterns are quite similar to those discussed in previous Section 1.6.1; for type L median outcomes are slightly more favourable than the mean values in column (1) of previous Table 1.10, because each school has the same weight in the computation of the median value across schools, although they have different share of students with low educated parents.[37]

The remaining columns of Table 1.11 give us a sense of how different those figures are across schools. Each column contains the difference between the expected outcome at the (100-p) and at the p percentile, with $p \in 20, 25, 30$. Here I will comment the results using the interquantile range (i.e. $p = 25$). Given the linear model for cognitive skills, differences in evaluations across schools are quite similar for the two types.[38] Differences are important: during the first level, students acquire cognitive skills 0.53 s.d. larger in the school at the 75 percentile than in the school at the 25 percentile (045 s.d. larger in the second level).

For type L, variation in graduation prospects across schools is quite large: the interquantile range is almost 19 p.p. This reflects sizable differences in dropout rate across schools. Moreover the large differences in retention rate may have important second order

[37] More specifically, for type L the median retention rate in first level is 45%, the median probability of dropping out immediately is more than 11% , that of not completing middle school is 24% , while chances of enrolling in high school are less that 33% (less than 45% conditional on graduation). Conversely, the median retention rate of type H is about 6% , less than 1% for dropout at $t = 1$, and less than 2% for not completing middle school; moreover the median rate of enrollment in high school is about 91% (93% conditional on graduation). In the median school the final level of cognitive skills among those who graduate are about -0.3 and 0.6 respectively.

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

effects on graduation, because retained students are more likely to leave school at time 1 and face the choice of dropout again at time 2. Interquantile ranges are definitely lower for type H: for instance it is 2 p.p. for graduation. In fact, variation of the underlying utilities across schools is similar for the two types, but given that their levels are much higher for type H, a unitary change in flow utility affects way less the probability of type H than the one of type L.

School environment also have a large effect on type L's probability of enrolling in high school (the interquantile range is about 19 p.p.). The effect is smaller but not negligible for type H (the interquantile range being 9 p.p.).

Tables 1.20 in the appendix shows that for the low type there would be large differences across schools even if peers were evenly distributed. Moreover difference in peers are associated as well with large variation in outcomes across school (Table 1.21). Peers and school effects do not move outcomes in the same direction: for a given outcome the correlation between average values simulated with constant peers or with constants school effects is typically low or negative.[39]

The rankings of schools based on the probability of dropout, graduation, or enrollment in high school are almost the same for the two types. However, for most of those events differences in probabilities across schools are so low for type H that moving from a top ranked school to one in the bottom tail would not dramatically change his prospects. Conversely, for type L the school environment can determine large changes in his expected attainments. Moreover the results of the simulation show that there is little correlation between schools' capability of increasing performances and their capability of leading everyone to graduation, or motivating students to choose further academic education. For instance, Figure 1.6 plots for each school the probability of graduation of type L (y-axis) and his predicted external cognitive skills at the end of middle school (x-axis). The two measures are positively correlated (0.4), in fact cognitive skills is one of the determinant of educational decisions. However there is clearly a lot of dispersion: many schools with average predicted cognitive skills have higher rates of graduation than schools with much larger predicted skills. Figure 1.7 shows a similar analysis for the type H student. While there is a similar large variation in expected cognitive skills across schools, the probability of graduation is above 90% in all the schools, and close to 1 in all schools with above average expected cognitive skills.

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

1.6.3 Retention and its consequences

Results in Section 2.4.1 show that repeating a level has a positive effect on cognitive skills accumulation, but it can also have a direct negative effect on educational choices, increasing probability of dropout and decreasing the probability of enrolling in high school, for a given level of cognitive skills. Moreover, as discussed in Section 1.5.2, both peers and school effects can impact the probability of retention. I now quantify the effect of repeating first or second level, comparing outcomes of students with identical characteristics, but different retention history.

Table 1.12 presents result for the type L student, using the average school environment among students with low educated parents as in column (1) of Table 1.10.[40] Retention at time 1 increases cognitive skills at the end of first level by about 0.2 s.d. Retention at time 2 for students enrolled in second level increases cognitive skills at the end of second level by about 0.4 s.d. Despite these positive effects on cognitive skills, retained students are by far less likely to complete middle school and enroll in high school. In particular, the ex-ante probability of graduation for a student retained at time $t = 1$ is about 56%, while for an identical student regularly promoted to next level it is 87%; moreover the ex-ante probability of enrolling in high school for the retained student is less than 13%, while it is about 42% for an identical student who did not experience retention. Graduation rate at time $t = 2$ for students who were not previously retained is 86%; students who fail at time 2 have more than 16% probability of dropout, therefore, despite having very high probability of graduating if they stay in school next period (almost 93%), they have a lower graduation rate of 78%. Moreover their probability of enrolling in high school is only 30% despite the large increase in cognitive skills.

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

Second, as shown in Table 1.12 which compare real and *perceived* cognitive skills, retained students have lower beliefs than identical students who were promoted: at time $t = 1$ the true level of cognitive skills for type L is -0.57, but the average perceived value for a retained student is lower (-0.61), while it is higher for a promoted student (-0.53). Similarly at time $t = 2$, the true value for a student in second level is -0.5, but on average student who graduate have a perceived value of -0.45 while those who are retained in second period on average believe that it is -0.57. Thus retained students are more likely to underestimate their true ability, and given that choices are based on beliefs, they would be more likely to dropout even in the absence of any direct negative effect of retention on

[40] To create the table I used the output of the baseline simulation discussed in previous Subsection 1.6.1, averaging outcomes by retention status.

their utility. On the other hand, students who complete middle school at time $t = 3$ have both actual and perceived level of cognitive skills higher than those of identical students who graduated at time $t = 2$, thus the large gaps in enrollment rate are surely due to differences in preferences.

To understand how important the uncertainty on own ability is in explaining the different choices of retained and regular students, I replicate analysis in Table 1.12 under the conterfactual scenario in which students' ability h is perfectly known rather than unobserved. Results in Table 1.17 show that removing uncertainty about ability would reduce dropout rate at the end of the period in which the student is retained by 1-2 p.p., increasing by a similar amount the probability of completing middle school. It would also slightly decrease the probability of enrolling in high school conditional on graduation (from about 50% to about 48%) for students who complete middle school at time $t = 2$, given that on average they slightly overestimate their cognitive skills in the baseline simulation.[41] Overall, the comparison of outcomes with and without uncertainty on ability shows that retained students are somewhat penalized by the randomness of the signals, but most of the differences in choices and attainments is due to changes in preferences.

Finally, Table 1.19 presents a drill down by retention status analogous to the one in Table 1.12, but using type H student. Outcomes for type H follows a similar pattern than those for type L. In fact, although he still has much better prospects than type L even when he repeats, retention in first level is associated with a 10 p.p. drop in his probability of graduating, while retention in second level with a 5 p.p. drop. Moreover his probability of enrolling in high school after graduation is largely impacted, because it shrinks from 95% to 77% (retention in first level) or 79% (retention in second level). It is, however, worth to recall that retention is a concern for a relatively small fraction of type H students (about 4% of them are retained in first level and only 1% in second level).

1.7 Counterfactual improvements of school effects

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

These counterfactual simulations can be regarded as government interventions target-

[41] For comparison, Table 1.18 replicates the analysis in Table 1.10 without uncertainty. Average outcomes are close enough with and without uncertainty.

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

1.7.1 Simulations using the entire sample

Table 1.13 summarizes outcomes under five counterfactual simulations; in all of them the targeted school effects are raised at the 75 percentile value if they are lower. The column "baseline" contains average outcomes under the benchmark model. In column $(\mathcal{A}_{\text{I}}, \mathcal{A}_{\text{II}})$ school effects on cognitive skills are modified. Column $(\mathcal{T}_{\text{I}})$ simulates an intervention on school effects on choice of staying in school, while in column $(\mathcal{T}_{\text{II}})$ school effects on choice of high school are improved; then in $(\mathcal{T}_{\text{I}}, \mathcal{T}_{\text{II}})$ both interventions are simultaneously implemented. Finally, in column $(\mathcal{R}_{\text{I}})$ school effects on retention at time $t = 1$ are adjusted. The table shows graduation rate, rate of enrollment in high school, average level of cognitive skils, and other outcomes under the baseline and the counterfactual scenario.[42]

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

By construction, interventions on "tastes" for education do not change average cognitive skills of the overall sample, and they can only decrease the average among graduated students if they avoid dropout of the less able students. In fact, a potential concern of implementing interventions aimed at improving other aspects than cognitive skills is that

[42]Note that in each counterfactual up to a quarter of schools (those with school effects above the 75 percentile) are not interested by the intervention, therefore their outcomes do not change. Computing average outcomes on the entire sample rather than on the subset of schools affected by each intervention makes results comparable across simulations.

they might keep in school or prompt to enroll in additional education students who do not have the necessary competences for it. Results in Table 1.13 suggest that this should not be a concern: interventions which raises $\mathcal{T}_{\mathrm{I}}$ or $\mathcal{T}_{\mathrm{II}}$ only cause a negligible decrease in the average cognitive skills after graduation, although they are associated with sizable increase in the graduation rate. More specifically, improving $\mathcal{T}_{\mathrm{I}}$ raising the lower values at the 75 percentile reduces dropout rate by 2 p.p. at time $t = 1$ and more than 3 p.p. at time $t = 2$. Graduation probabilities among this larger pool are almost unchanged. Enrollment in high school is also almost the same. Improving school effects on the choice of enrolling in upper secondary education ($\mathcal{T}_{\mathrm{II}}$) reduce dropout during middle school (of almost 1 p.p. at time 1 and almost 2 p.p. at time 2). Here the channel to increase graduation rate is only the higher future utility from high school attendance. Enrollment in high school is raised by 4 p.p.

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

The last column of the table shows average outcomes when school effects $\mathcal{R}_{\mathrm{I}}$ on retention are modified to make schools more lenient. This intervention reduces the share of retained students by almost 4 p.p.; the second order effects are a 1 p.p. increase in graduation rate. The overall enrollment in high school increases as well by 1 p.p.[43]

Graduation rate and enrollment in high school increase even more for the subsample of students with low parental background, as shown in Table 1.23 in the appendix. The table uses the same outputs of Table 1.13, but frequencies are computed using only individuals with low educated parents. Raising school effects increases graduation rate by up to 6 p.p. and enrollment in high school by up to 6.5 p.p. Next subsection focuses on studying low parental background students and how they are affected by the interventions depending on their school environment.

1.7.2 Simulations on type L students

I now study what outcomes students with low parental background would have under the various interventions described in previous section. In particular, I analyze how the interventions affect students enrolled in schools with a high or with a low share of other low parental background students. As described in Section 1.6.1, schools with high share of students with low educated parents are more likely to have higher school effect on choice

[43]For comparison Table 1.22 in the appendix replicates the exercise described in this section using the median rather than the 75 percentile as threshold. Not surprisingly differences across baseline and counterfactuals are smaller, but they follow a similar pattern. The main difference is that the intervention which raises both $\mathcal{T}_{\mathrm{I}}$ and $\mathcal{T}_{\mathrm{II}}$ increases graduation and enrollment in high school more than the intervention which raises $\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$

of staying in school, and slightly more likely to have lower school effect on cognitive skills, but there is large variation in all dimensions. Obviously students enrolled in a school with very large school effects on choices would not be affected by an intervention that raises them, while they may benefit from higher school effects on cognitive skills, and vice-versa for students enrolled in schools with high school effects on cognitive skills.

I simulated outcomes for type L in each school in the sample under the various counterfactuals, following the steps described in Section 1.6. Table 1.14 shows predicted outcomes for individuals enrolled in schools with high share of students with low educated parents (panel A), and for individuals enrolled in schools with low share of students with low educated parents (panel B).[44]

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

In fact, on average raising school effects on cognitive skills at the 75 percentile improves $C_{\mathrm{I},1}$ by 0.26 s.d. in schools with high share of low parental background students, and by 0.12 s.d. in schools with low share. In the former group of schools, both graduation rate and enrollment in high school increase by 12 p.p., while raising simultaneously the two school effects on tastes increases graduation rate by about 4 p.p. and enrollment in high school by 5.5 p.p. In the latter group of schools, raising $\mathcal{A}_{\mathrm{I}}$ and $\mathcal{A}_{\mathrm{II}}$ increases graduation rate by 5 p.p. and enrollment in high school by 7.5 p.p., while raising $\mathcal{T}_{\mathrm{I}}$ and $\mathcal{T}_{\mathrm{II}}$ increases graduation rate by 9.5 p.p. and enrollment in high school by 12 p.p.

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

1.8 Conclusions and future research

Suppose that a policy maker wants to identify the "most successful" institutions, in order to investigate their methodology, learn their best practices, and transfer them to other schools. The school effects on cognitive skills identified through the model presented in this paper or other similar measures of value added would allow her to rank schools based on their capability to improve performance as measured by a nation-wide test. However, it is not evident that attending one of the top performing institutions would be desirable

[44] I ranked schools based on their share of students whose parents attained at most lower secondary education, and grouped them in three quantiles. The 15 schools in the top quantile are used for panel A of Table 1.14, the 15 schools in the bottom quantile are used for panel B.

for every type of student, if those institutions do not simultaneously ensure that they help every student to succeed. In fact, graduating from such schools students would potentially reach the highest level of cognitive skills, but this is not happening in practice if they dropout before completing their education. Moreover, in another school they may graduate with a slightly lower level of cognitive skills, but with a stronger motivation to enroll in high school, which may eventually lead to better outcomes in the labor market.

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

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

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

1.9 Tables

Table 1.1: Descriptive statistics by subgroups of the population

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

Note. The table reports summary statistics for the sample of students used to estimate my structural model. It consists of students who enrolled in a public middle school in Barcelona (Spain) in 2009 or in 2010. eval. PS are external evaluations at the end of primary school; eval. MS are external evaluations at the end of middle school (computed on the subsample who reached the last grade).

Table 1.2: Descriptive statistics by schools

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

Note. This table reports summary statistics for the 44 public middle schools in Barcelona which are used to estimate the structural model discussed in this paper.

Table 1.3: Variance of unobserved ability

	$\mu_l^2\widehat{\sigma}$	$\text{Var}(C_l)$	%
$l = 0$	0.277 (0.012)	0.563 (0.016)	49.3 (1.5)
$l = \text{I}$	0.592 (0.017)	1.201 (0.048)	49.3 (1.5)
$l = \text{II}$	0.612 (0.019)	1.045 (0.039)	58.6 (1.3)

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

Time	Signals received from 0 to t	Posterior Variance
$t = 0$	r_0	0.1693 (0.0041)
$t = 1$	$r_0, g_{\text{I},1}$	0.0600 (0.0029)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{I},2}$	0.0364 (0.0020)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{II},2}$	0.0349 (0.0021)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}$	0.0293 (0.0015)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{I},2}, g_{\text{II},3}$	0.0254 (0.0015)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{I},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0223 (0.0012)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, g_{\text{II},3}$	0.0246 (0.0016)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}, g_{\text{II},3}$	0.0217 (0.0013)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0217 (0.0013)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0194 (0.0010)

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

Table 1.4: Estimates of evaluations parameters. Periods 0 and I

	r_0	g_{I}		
	β_0	β_{I}	γ	$\beta_{\mathrm{I}} + \gamma$
Female	0.044 (0.023)	-0.047 (0.108)	0.372 (0.041)	0.326 (0.097)
Immigrant	-0.263 (0.036)	-0.159 (0.205)	0.091 (0.066)	-0.068 (0.170)
Mother edu. avg	0.233 (0.031)	0.325 (0.132)	-0.090 (0.053)	0.235 (0.104)
Mother edu. high	0.472 (0.038)	0.686 (0.122)	-0.108 (0.072)	0.578 (0.102)
Father education average	0.236 (0.029)	0.198 (0.125)	-0.012 (0.041)	0.186 (0.114)
Father education high	0.379 (0.034)	0.232 (0.126)	0.007 (0.051)	0.239 (0.114)
Day of birth	0.246 (0.048)	0.138 (0.060)	0.006 (0.044)	0.144 (0.042)
Retained in primary school	-0.556 (0.041)	-0.704 (0.096)	0.379 (0.070)	-0.325 (0.049)
Repeat level		0.166 (0.063)	-0.181 (0.051)	-0.015 (0.032)
Peers: avg evaluation PS		0.175 (0.059)	-0.136 (0.039)	0.039 (0.042)
Peers: avg parental edu		0.128 (0.072)	-0.076 (0.047)	0.053 (0.042)
Peers: share of female		0.063 (0.020)	-0.024 (0.012)	0.038 (0.014)
Peers: share of immigrant		-0.016 (0.040)	-0.031 (0.032)	-0.047 (0.029)
Peers: share with external		0.013 (0.041)	0.014 (0.028)	0.027 (0.021)
Peers: share older		0.000 (0.040)	0.011 (0.025)	0.011 (0.023)
School effect p75 - p25		0.312 (0.092)	0.395 (0.086)	0.255 (0.055)
School effect p80 - p20		0.416 (0.102)	0.503 (0.088)	0.280 (0.065)
School effect st. dev.		0.266 (0.054)	0.268 (0.049)	0.178 (0.027)
Female X school eff.		-0.016 (0.051)	-0.003 (0.033)	
Immigrant X school eff.		-0.051 (0.088)	-0.024 (0.051)	
Mother edu. avg. X school eff.		-0.030 (0.056)	-0.032 (0.048)	
Mother edu. high X school eff.		-0.025 (0.050)	-0.023 (0.040)	
Father edu. avg X school eff.		0.029 (0.056)	0.036 (0.033)	
Father edu. high X school eff.		0.079 (0.047)	0.027 (0.050)	
Previous time-varying regressors		0.058 (0.060)		0.058 (0.060)
Unobserved ability	1	1.461 (0.035)		
Variance of error	0.434 (0.011)			0.198 (0.008)

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

Bootstrap standard errors in parentheses.

Table 1.5: Estimates of evaluations parameters. Period II

	r_{II}	g_{II}	
	β_{II}	$\mu\beta_{\text{II}}$	$\mu\beta_{\text{II}} + \gamma$
Female	0.095 (0.095)	0.111 (0.111)	0.484 (0.101)
Immigrant	-0.140 (0.157)	-0.165 (0.187)	-0.074 (0.169)
Mother edu. avg	0.266 (0.090)	0.313 (0.105)	0.223 (0.094)
Mother edu. high	0.533 (0.079)	0.627 (0.093)	0.519 (0.102)
Father education average	0.233 (0.100)	0.274 (0.120)	0.262 (0.116)
Father education high	0.262 (0.095)	0.308 (0.116)	0.315 (0.118)
Day of birth	0.115 (0.052)	0.136 (0.062)	0.142 (0.051)
Retained in primary school	-0.656 (0.068)	-0.772 (0.074)	-0.393 (0.045)
Repeat level	0.343 (0.051)	0.403 (0.059)	0.222 (0.042)
Peers: avg evaluation PS	0.082 (0.027)	0.096 (0.032)	-0.040 (0.028)
Peers: avg parental edu	0.026 (0.026)	0.031 (0.031)	-0.045 (0.032)
Peers: share of female	-0.003 (0.009)	-0.004 (0.011)	-0.028 (0.011)
Peers: share of immigrant	-0.021 (0.023)	-0.025 (0.027)	-0.055 (0.033)
Peers: share with external	-0.030 (0.018)	-0.036 (0.022)	-0.022 (0.022)
Peers: share older	0.011 (0.016)	0.012 (0.019)	0.023 (0.019)
School effect p75 - p25	0.241 (0.059)	0.284 (0.073)	0.285 (0.041)
School effect p80 - p20	0.296 (0.064)	0.348 (0.078)	0.346 (0.065)
School effect st. dev.	0.224 (0.035)	0.263 (0.043)	0.195 (0.029)
Female X school eff.	-0.041 (0.033)	-0.048 (0.039)	
Immigrant X school eff.	-0.041 (0.066)	-0.049 (0.079)	
Mother edu. avg. X school eff.	-0.029 (0.036)	-0.034 (0.042)	
Mother edu. high X school eff.	-0.003 (0.035)	-0.004 (0.041)	
Father edu. avg X school eff.	-0.007 (0.043)	-0.008 (0.052)	
Father edu. high X school eff.	0.030 (0.036)	0.035 (0.042)	
Previous time-varying regressors	0.443 (0.080)	0.521 (0.089)	0.521 (0.089)
Unobserved ability	1.262 (0.032)	1.486 (0.039)	
Variance of error	0.293 (0.009)		0.185 (0.009)

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

Bootstrap standard errors in parentheses.

Table 1.6: Estimates of retention and graduation parameters

	Retention in I	Graduation in II
Belief cognitive skills	-1.791 (0.105)	2.479 (0.192)
Female	-0.723 (0.091)	0.883 (0.161)
Immigrant	0.028 (0.130)	0.164 (0.249)
Mother education average	0.018 (0.134)	-0.107 (0.186)
Mother education high	-0.277 (0.157)	0.194 (0.201)
Father education average	-0.126 (0.127)	0.002 (0.175)
Father education high	-0.480 (0.188)	0.227 (0.246)
Day of birth	-0.082 (0.171)	0.130 (0.284)
Retained in primary school		0.569 (0.306)
Repeated I		0.179 (0.224)
Second time in II		0.005 (0.307)
Peers: avg evaluation PS	0.216 (0.240)	0.262 (0.259)
Peers: avg parental edu	0.033 (0.196)	-0.547 (0.246)
Peers: share of female	-0.215 (0.088)	-0.168 (0.099)
Peers: share of immigrant	0.029 (0.151)	-0.339 (0.170)
Peers: share with external	-1.055 (0.162)	-0.092 (0.185)
Peers: share older	0.748 (0.127)	0.091 (0.128)
School effect p75 - p25	0.654 (0.195)	1.378 (0.258)
School effect p80 - p20	1.083 (0.214)	1.865 (0.289)

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

Bootstrap standard errors in parentheses.

Table 1.7: Estimates of choices parameters

	Stay in middle school	Enroll in high school
$\widehat{C}$ (belief on cognitive skills)	1.412	2.019
$\widehat{C}\times$individual characteristics	0.209	-0.196
$\widehat{C}\times$school effect	0.040	-0.486
$\widehat{C}\times$peers characteristics	0.136	0.959
$\widehat{C}\times$second time in I	-0.407	
$\widehat{C}\times$second time in II	-0.334	-0.723
$\widehat{C}\times$in II after repeating I	-0.499	-0.578
Female	0.354	0.747
Immigrant	-0.399	0.611
Mother education average	-0.073	0.019
Mother education high	0.078	-0.182
Father education average	-0.225	0.291
Father education high	-0.473	0.623
Day of birth	-0.549	-0.155
Retained in primary school	0.439	0.956
Second time in I	-1.283	
Second time in II	-0.305	-1.628
In II after repeating I	-0.716	-1.716
Peers: avg evaluation PS	-0.333	0.023
Peers: avg parental edu	0.203	-0.106
Peers: share of female	-0.106	0.196
Peers: share of immigrant	-0.094	-0.205
Peers: share with external	0.224	0.103
Peers: share older	-0.034	0.149
School effect p75 - p25	0.668	0.630
School effect p80 - p20	0.913	0.689

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

Computation of standard errors under completion.

Table 1.8: 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	91.36	91.75	83.20	83.86	65.66	65.69	23.42	23.02	4.26	6.01
male	90.28	90.13	80.09	80.21	60.11	59.55	26.79	26.58	4.82	6.98
female	92.50	93.46	86.49	87.72	71.54	72.19	19.85	19.25	3.67	4.98
Spanish	93.90	94.02	86.81	87.59	69.54	69.34	18.53	18.26	4.11	5.93
immigrant	77.34	79.25	63.28	63.27	44.27	45.57	50.39	49.27	5.08	6.45
low parental edu.	83.90	84.00	69.83	68.56	42.29	41.38	43.47	43.12	6.86	8.31
avg parental edu.	90.19	91.27	81.08	82.69	62.01	62.71	25.70	24.85	4.81	7.19
high parental edu.	97.88	97.63	95.15	95.74	85.78	85.40	6.97	6.96	1.95	3.15
below median peers	86.82	88.25	76.21	77.58	54.00	54.94	32.78	31.69	4.95	6.22
above median peers	94.65	94.30	88.27	88.41	74.12	73.48	16.63	16.73	3.76	5.86

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

Table 1.9: Average school environment by student type

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

Panel A: Peers' characteristics

	$\mathcal{I}$	$\mathcal{A}_{\text{I}}$	$\mathcal{A}_{\text{II}}$	$\mathcal{T}_{\text{I}}$	$\mathcal{T}_{\text{II}}$	$\mathcal{R}_{\text{I}}$	$\mathcal{G}_{\text{II}}$
low parental edu.	-0.196	0.608	0.624	-0.496	-0.191	0.415	-0.496
(percentile)	54	46	48	55	42	42	53
high parental edu.	-0.239	0.672	0.690	-0.814	-0.228	0.704	-0.412
(percentile)	44	46	52	37	39	57	54

Panel B: School effects

Table 1.10: Educational outcomes by student type and environment

	Avg. p.e.	Low parental education				High parental education			
	(1)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
repeat level I	0.282	0.596	0.632	0.302	0.337	0.039	0.034	0.132	0.117
dropout at t=1	0.054	0.144	0.197	0.065	0.090	0.007	0.005	0.022	0.017
drop at t=2|not grad.	0.143	0.232	0.275	0.153	0.187	0.050	0.042	0.098	0.080
graduate	0.892	0.687	0.631	0.836	0.809	0.991	0.992	0.965	0.973
enrol in high school	0.652	0.249	0.239	0.433	0.434	0.938	0.933	0.869	0.869
enrol in h.s.|grad.	0.731	0.362	0.379	0.518	0.537	0.947	0.941	0.901	0.893
r_0	0.064	-0.349	-0.349	-0.349	-0.349	0.502	0.502	0.502	0.502
$C_{\mathrm{I},1}$	0.107	-0.577	-0.513	-0.257	-0.193	0.860	0.783	0.541	0.464
$g_{\mathrm{I},1}$	-0.039	-0.472	-0.450	-0.427	-0.406	0.543	0.509	0.498	0.465
r_{II}	0.113	-0.379	-0.326	-0.187	-0.138	0.697	0.633	0.510	0.446
g_{II}	-0.124	-0.458	-0.435	-0.491	-0.471	0.422	0.390	0.451	0.418

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

Table 1.11: Educational outcomes by schools

	Low parental education				High parental education			
	p50	p80-p20	p75-p25	p70-p30	p50	p80-p20	p75-p25	p70-p30
repeat level I	0.454	0.314	0.241	0.190	0.063	0.090	0.071	0.049
dropout at t=1	0.111	0.106	0.087	0.055	0.008	0.019	0.013	0.009
drop. at t=2|do not grad.	0.203	0.108	0.082	0.068	0.074	0.045	0.034	0.025
graduate	0.762	0.204	0.185	0.137	0.988	0.027	0.023	0.015
enrol in high school	0.327	0.253	0.187	0.152	0.909	0.103	0.092	0.063
enrol in h.s.|graduate	0.445	0.248	0.194	0.141	0.918	0.099	0.063	0.048
$C_{\mathrm{I},1}$	-0.374	0.636	0.532	0.455	0.664	0.708	0.587	0.469
C_{II}|graduate	-0.302	0.565	0.455	0.381	0.637	0.623	0.483	0.428

Note. Frequencies are constructed using 10000 simulations of the structural model for each type and every school of the sample. Type L has parents with primary education; type H has parents with tertiary education.The fictitious students created in the simulation are male, Spanish, born on July 1, they began middle school at 12 years old and they are assigned the average cohort effect and the average primary school effect in the sample. The average peers characteristics in each school are used. The first column of each part of the table contains the median outcome, the other columns contain difference between quantiles.

Table 1.12: Educational outcomes by retention status. Student with low educated parents

	not retained	retained in I	retained in II
dropout at t=1	0.0992	0.1752	-
graduate (Pr at t=1)	0.8735	0.5610	-
enrol in high school (Pr at t=1)	0.4233	0.1306	-
graduate at t=2|stay at t=1	0.8618	-	-
enrol in hs at t=2|grad. at t=2	0.5079	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.2391	0.1579
graduate at t=3|stay at t=2	-	0.8939	0.9271
enrol in hs at t=3|grad. at t=3	-	0.2329	0.2982
true $C_{\text{I},1}$	-0.5765	-0.5765	-
perceived $\widehat{C}_{\text{I},1}$	-0.5281	-0.6086	-
true $C_{\text{I},2}$	-	-0.4102	-
perceived $\widehat{C}_{\text{I},2}$	-	-0.4106	-
true $C_{\text{II},2}$	-0.5017	-	-0.5017
perceived $\widehat{C}_{\text{II},2}$	-0.4546	-	-0.5608
true C_{II}|graduate	-0.5017	-0.4150	-0.0985
perceived $\widehat{C}_{\text{II}}$|graduate	-0.4546	-0.3920	-0.1229

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

Table 1.13: Simulated outcomes for the entire sample. School effects raised at 75 percentile.

	Baseline	Counterfactuals				
school effects modified		$\mathcal{A}_{\text{I}}, \mathcal{A}_{\text{II}}$	$\mathcal{T}_{\text{I}}$	$\mathcal{T}_{\text{II}}$	$\mathcal{T}_{\text{I}}, \mathcal{T}_{\text{II}}$	$\mathcal{R}_{\text{I}}$
repeat level I	0.230	0.205	0.230	0.230	0.230	0.191
dropout at t=1	0.083	0.065	0.062	0.077	0.057	0.079
grad at t=2| in II	0.921	0.936	0.918	0.920	0.917	0.915
drop. at t=2|not grad.	0.212	0.191	0.178	0.195	0.165	0.215
grad at t=3| in II	0.850	0.867	0.833	0.849	0.832	0.843
graduate	0.839	0.871	0.859	0.847	0.865	0.849
enrol in high school	0.659	0.708	0.668	0.701	0.711	0.670
enrol in hs|graduation	0.785	0.812	0.778	0.827	0.822	0.789
$C_{\text{I},1}$	-0.000	0.173	-0.000	-0.000	-0.000	-0.000
C_{II}|graduation	0.272	0.390	0.256	0.265	0.250	0.265

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

Table 1.14: Simulated outcomes for type L. School effects raised at 75 percentile.

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

	Baseline	Counterfactuals				
school effects modified		$\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}$	$\mathcal{T}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}}$	$\mathcal{R}_{\mathrm{I}}$
repeat level I	0.573	0.478	0.573	0.573	0.573	0.502
dropout at t=1	0.163	0.102	0.137	0.155	0.130	0.156
grad at t=2| in II	0.830	0.886	0.828	0.830	0.828	0.810
drop. at t=2|not grad.	0.262	0.212	0.234	0.253	0.226	0.263
grad at t=3| in II	0.858	0.912	0.855	0.859	0.856	0.863
graduate	0.645	0.765	0.675	0.656	0.685	0.666
enrol in high school	0.226	0.346	0.235	0.269	0.281	0.242
enrol in hs|graduation	0.351	0.453	0.348	0.410	0.410	0.364
$C_{\mathrm{I},1}$	-0.793	-0.532	-0.793	-0.793	-0.793	-0.793
C_{II}|graduation	-0.581	-0.373	-0.573	-0.579	-0.572	-0.574

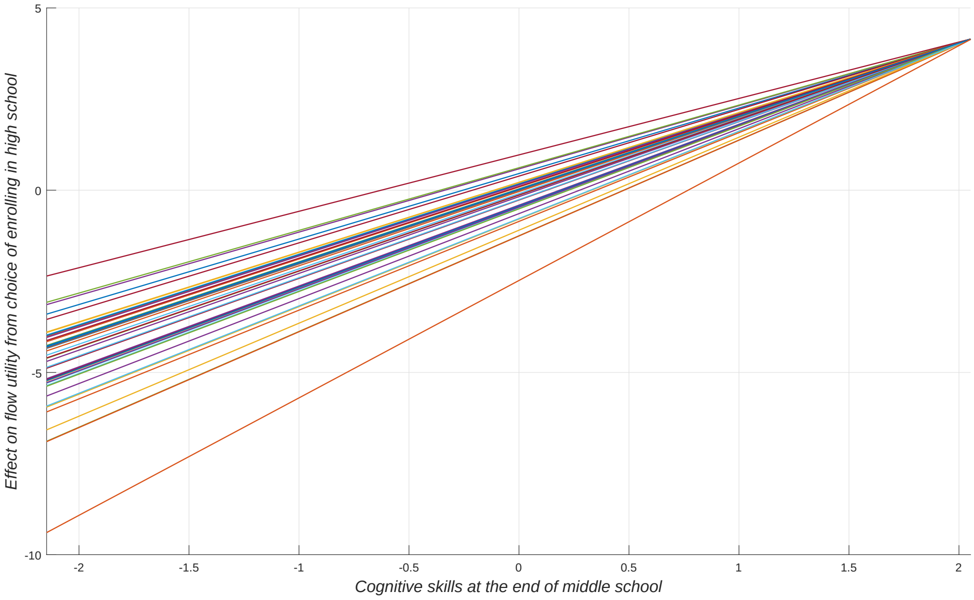
Panel B: Schools with low share of students with low educated parents

	Baseline	Counterfactuals				
school effects modified		$\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}$	$\mathcal{T}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}}$	$\mathcal{R}_{\mathrm{I}}$
repeat level I	0.356	0.316	0.356	0.356	0.356	0.215
dropout at t=1	0.101	0.076	0.039	0.088	0.035	0.085
grad at t=2| in II	0.843	0.874	0.840	0.842	0.839	0.833
drop. at t=2|not grad.	0.229	0.205	0.136	0.205	0.120	0.217
grad at t=3| in II	0.894	0.912	0.893	0.896	0.893	0.896
graduate	0.777	0.827	0.861	0.797	0.871	0.825
enrol in high school	0.446	0.521	0.481	0.527	0.565	0.502
enrol in hs|graduation	0.574	0.630	0.559	0.662	0.649	0.608
$C_{\mathrm{I},1}$	-0.110	0.013	-0.110	-0.110	-0.110	-0.110
C_{II}|graduation	-0.061	0.046	-0.050	-0.054	-0.048	-0.054

Note. Average outcomes in the column “baseline” are computed using the estimated parameters of the model. In the counterfactuals, the school effects reported above each column are replaced with values at the 75 percentile if they are lower. Frequencies are constructed using 10000 simulations of the structural model for each school of the type L student described in Section 1.6. Panel A shows average outcomes among individual enrolled in schools with high share of students with low educated parents (the top third of schools); panel B shows average outcomes for schools with low share of students with low educated parents (the bottom third).

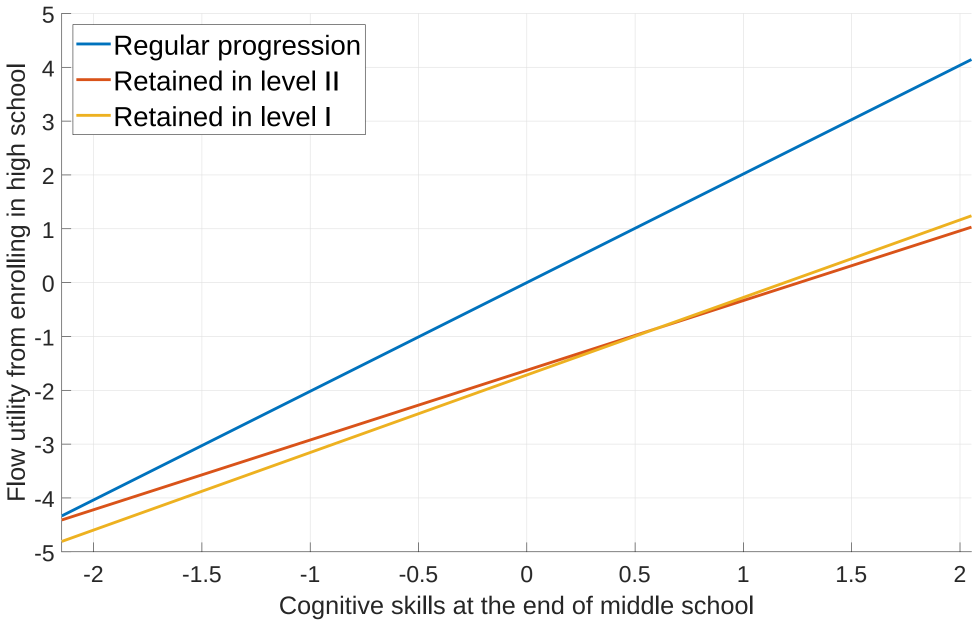
1.10 Figures

Figure 1.1: Effect of cognitive skills on high school enrollment by middle school attended



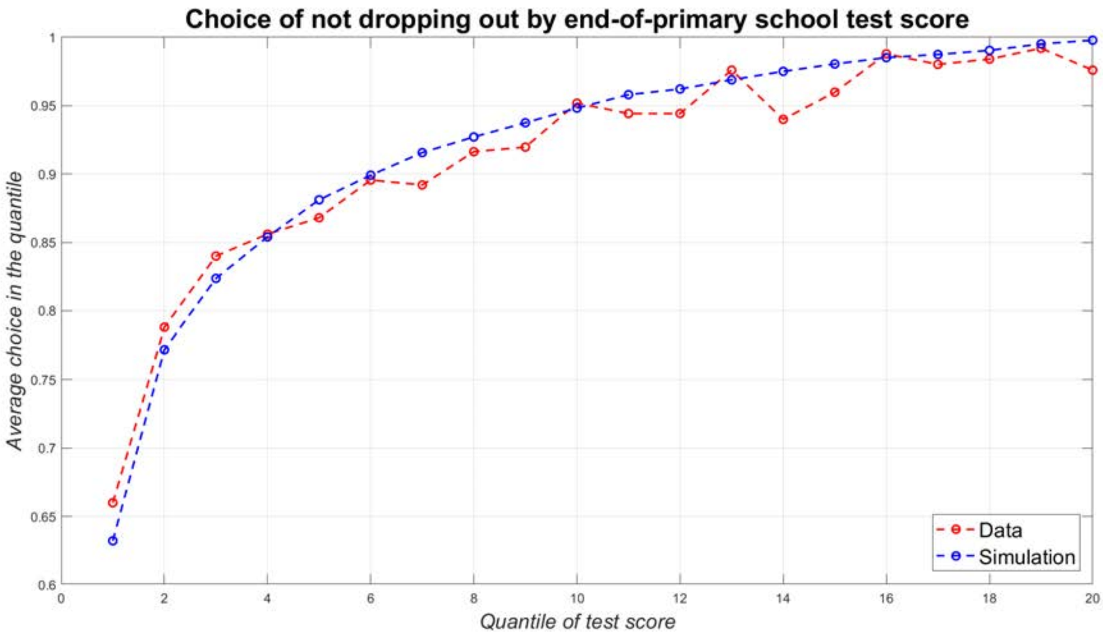
Note. The figure plots the effect of beliefs about cognitive skills on the flow utility of the choice of enrolling in high school after graduation. Each line represents a school in the sample.

Figure 1.2: Effect of cognitive skills on high school enrollment by retention status



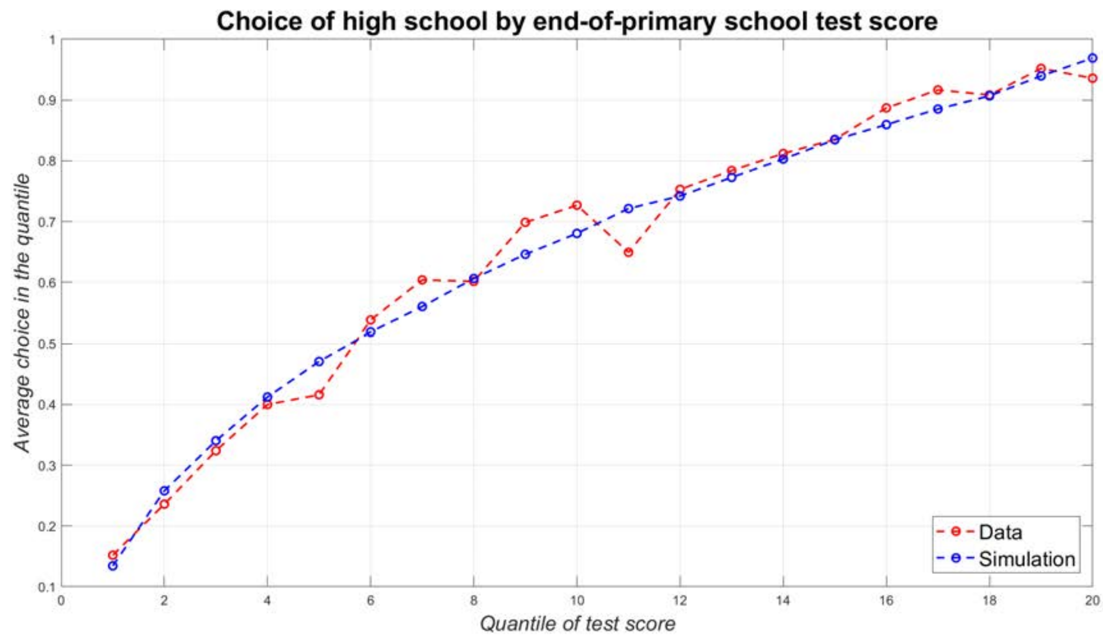
Note. The figure plots the effect of beliefs about cognitive skills on the utility from the choice of enrolling in high school for three types of students: those who completed middle school regularly at time $t = 2$ (blue line), those who graduated at time $t = 3$ because they were retained at time$t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

Figure 1.3: Fit of the model (i)



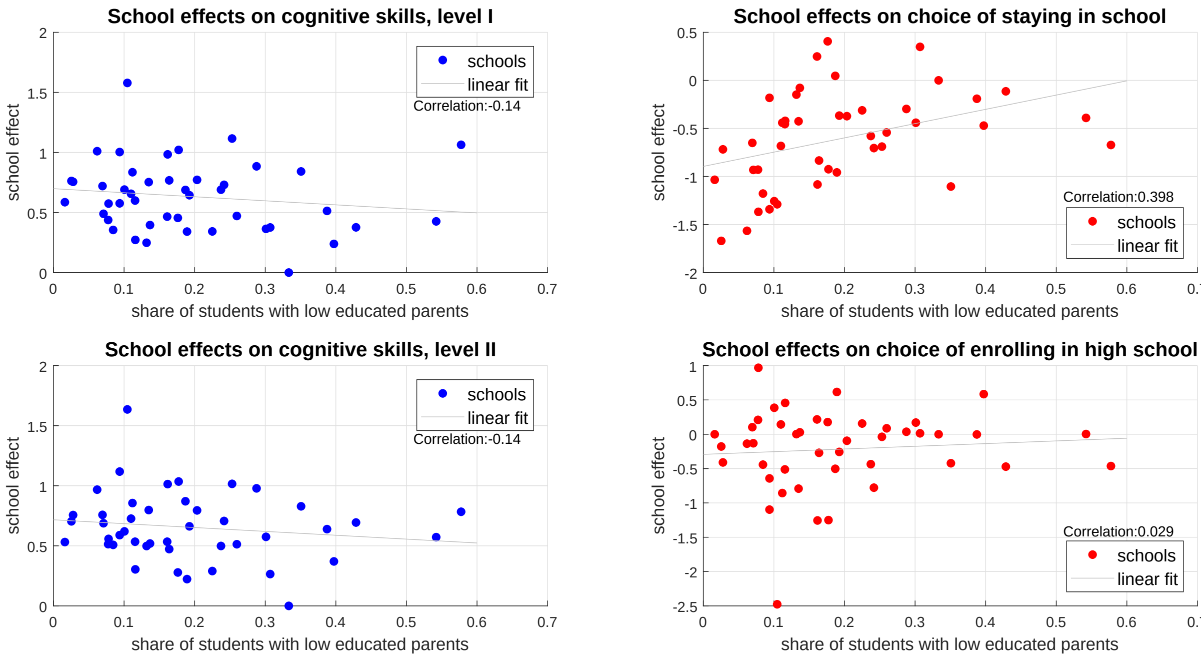
Note. The figure plots the share of students who chose to stay in school at time $t = 1$ by quantile of their test score at the end of primary school. Sample average from the real data are in red, while results of the simulation performed using the estimated parameters of the model are in blue.

Figure 1.4: Fit of the model (ii)



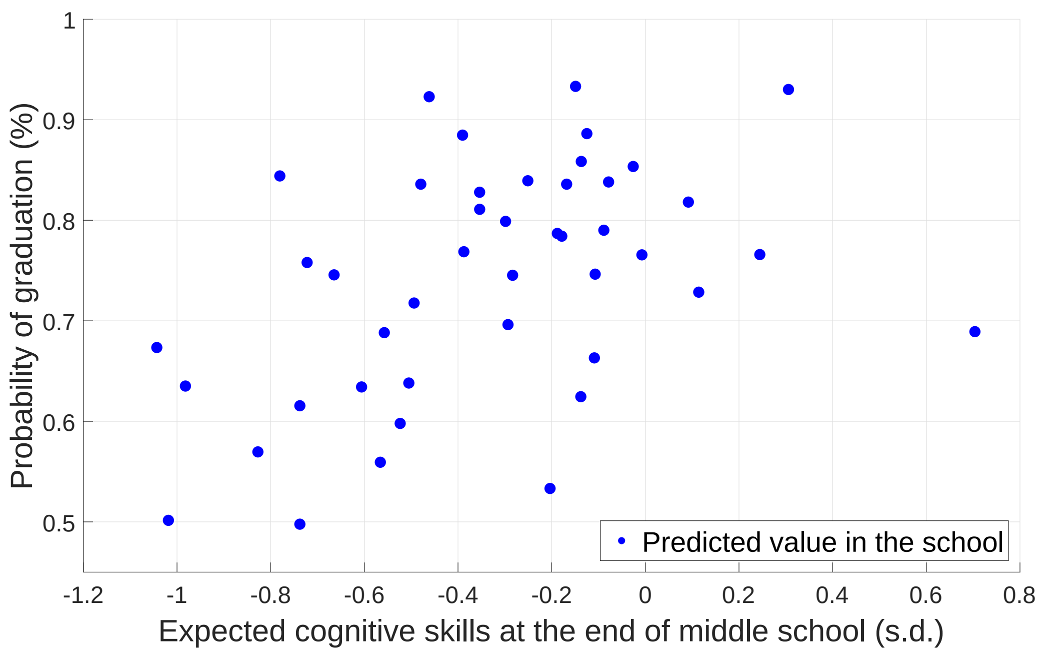
Note. The figure plots the share of students who enroll in high school by quantile of their test score at the end of primary school. Sample average from the real data are in red, while results of the simulation performed using the estimated parameters of the model are in blue.

Figure 1.5: Share of students with low educated parents *vs* school effects



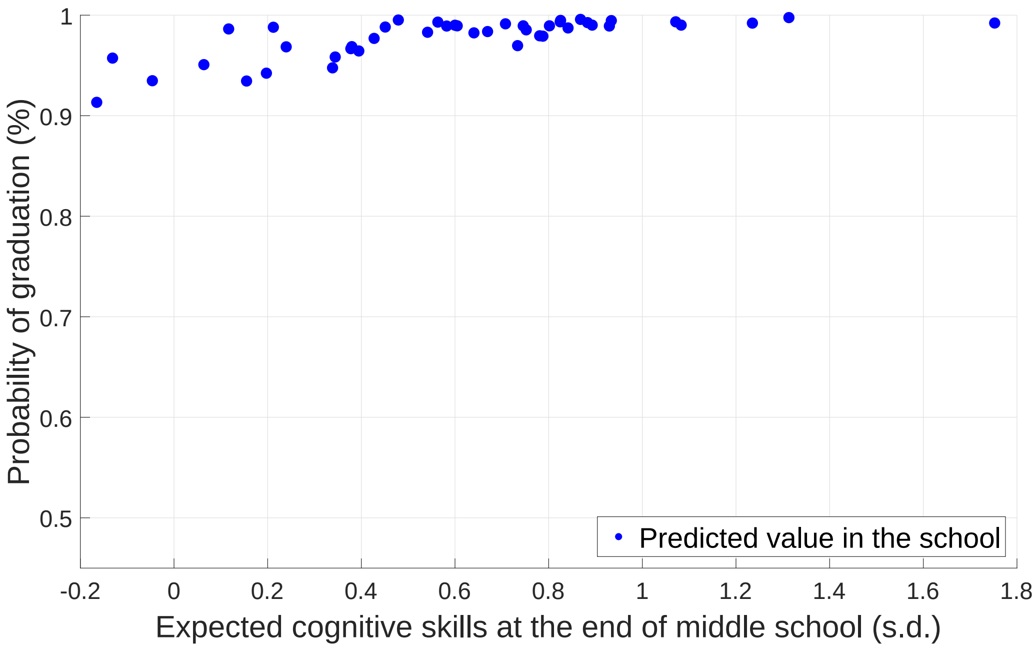
Note. The figure plots the estimated school effects against the share of students with low educated parents (at most lower secondary education) enrolled in the school. The left side plots school effects on cognitive skills, the right side plots school effects on choices.

Figure 1.6: Student with low educated parents: expected outcomes by school



Note. The figure plots simulated outcomes for a Spanish male student of average innate ability whose parents are low educated (at most lower secondary education). Expected outcomes at the school level are computed using data of the simulation described in Section 1.6.2 and Table 1.11.

Figure 1.7: Student with highly educated parents: expected outcomes by school



Note. The figure plots simulated outcomes for a Spanish male student of average innate ability whose parents have tertiary education. Expected outcomes at the school level are computed using data of the simulation described in Section 1.6.2 and Table 1.11.

Appendix 1.A Additional tables

Table 1.15: Estimates of evaluations parameters. Coefficients peers and interactions

	β_{I}	β_{II}	γ
Peers: share of female	0.040 (0.016)	-0.000 (0.007)	-0.020 (0.010)
	0.003 (0.017)	-0.005 (0.006)	0.012 (0.007)
	-0.032 (0.012)	-0.000 (0.006)	0.021 (0.008)
Peers: share of immigrant	-0.002 (0.034)	-0.010 (0.017)	0.004 (0.024)
	-0.022 (0.021)	-0.015 (0.011)	0.010 (0.013)
	0.018 (0.022)	0.013 (0.010)	0.025 (0.017)
Peers: avg parental edu	0.173 (0.056)	0.020 (0.026)	-0.101 (0.038)
	-0.025 (0.033)	0.009 (0.013)	0.016 (0.020)
	0.048 (0.021)	0.021 (0.008)	-0.022 (0.013)
Peers: share with external	0.004 (0.028)	0.016 (0.012)	-0.005 (0.023)
	-0.003 (0.022)	-0.021 (0.011)	0.013 (0.018)
	0.010 (0.022)	-0.011 (0.012)	-0.000 (0.021)
Peers: avg evaluation PS	0.154 (0.037)	0.068 (0.022)	-0.181 (0.029)
	0.019 (0.025)	0.009 (0.013)	0.015 (0.019)
	-0.004 (0.025)	0.001 (0.016)	0.022 (0.019)
Peers: share older	-0.001 (0.036)	0.010 (0.013)	0.012 (0.022)
	0.028 (0.017)	0.008 (0.008)	0.001 (0.010)
	-0.012 (0.011)	-0.003 (0.007)	0.001 (0.008)
Female X school eff.	-0.060 (0.202)	-0.184 (0.146)	-0.012 (0.117)
Immigrant X school eff.	-0.191 (0.322)	-0.185 (0.282)	-0.090 (0.181)
Mother edu. avg. X school eff.	-0.114 (0.214)	-0.130 (0.170)	-0.118 (0.172)
Mother edu. high X school eff.	-0.094 (0.171)	-0.014 (0.149)	-0.084 (0.133)
Father edu. avg X school eff.	0.107 (0.232)	-0.032 (0.192)	0.135 (0.129)
Father edu. high X school eff.	0.296 (0.263)	0.135 (0.161)	0.100 (0.200)

Note. See Tables 1.4 and 1.5.

Table 1.16: Fit of the model (bis)

	graduate at $t = 2$		graduate at $t = 3$		enroll in hs	
	Data	Model	Data	Model	Data	Model
ALL	94.28	91.86	81.40	85.07	78.92	78.33
male	93.23	90.02	78.62	82.64	75.05	74.24
female	95.29	93.62	84.94	88.28	82.71	82.29
Spanish	94.87	92.49	81.57	85.90	80.10	79.16
immigrant	88.15	85.96	80.99	82.38	69.96	72.03
low parental edu.	87.20	84.07	83.18	81.64	60.56	60.36
avg parental edu.	93.26	89.96	80.50	85.71	76.48	75.84
high parental edu.	97.89	96.55	82.52	92.64	90.15	89.19
below median peers	92.26	90.32	82.09	84.44	70.85	70.82
above median peers	95.41	92.76	80.56	85.75	83.97	83.11

Note. Statistics are conditional on reaching the relevant level

Table 1.17: Educational outcomes by retention status. Student with low educated parents. Known ability.

	not retained	retained in I	retained in II
dropout at t=1	0.1061	0.1597	-
graduate (Pr at t=1)	0.8705	0.5789	-
enrol in high school (Pr at t=1)	0.4032	0.1281	-
graduate at t=2|stay at t=1	0.8653	-	-
enrol in hs at t=2|grad. at t=2	0.4836	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.2383	0.1455
graduate at t=3|stay at t=2	-	0.9045	0.9430
enrol in hs at t=3|grad. at t=3	-	0.2212	0.3002
true $C_{\mathrm{I},1}$	-0.5765	-0.5765	-
true $C_{\mathrm{I},2}$	-	-0.4102	-
true $C_{\mathrm{II},2}$	-0.5017	-	-0.5017
true C_{II}|graduate	-0.5017	-0.4150	-0.0985

Table 1.18: Educational outcomes by student type and environment - with known ability

	Avg p.e.	Low parental education				High parental education			
	(1)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
repeat level I	0.266	0.604	0.646	0.286	0.323	0.033	0.028	0.114	0.100
dropout at t=1	0.043	0.138	0.193	0.059	0.082	0.004	0.003	0.014	0.011
drop at t=2|not grad.	0.133	0.230	0.278	0.144	0.179	0.041	0.033	0.086	0.070
graduate	0.911	0.694	0.631	0.855	0.828	0.994	0.995	0.976	0.981
enrol in high school	0.674	0.237	0.224	0.440	0.441	0.946	0.942	0.886	0.885
enrol in hs|graduation	0.740	0.341	0.356	0.514	0.533	0.952	0.947	0.909	0.902

Table 1.19: Educational outcomes by retention status. Student with high educated parents

	not retained	retained in I	retained in II
dropout at t=1	0.0057	0.0347	-
graduate (Pr at t=1)	0.9938	0.9102	-
enrol in high school (Pr at t=1)	0.9476	0.7011	-
graduate at t=2|stay at t=1	0.9901	-	-
enrol in hs at t=2|grad. at t=2	0.9550	-	-
dropout at t=2|stay at t=1 & do not grad.	-	0.0517	0.0434
graduate at t=3|stay at t=2	-	0.9945	0.9956
enrol in hs at t=3|grad. at t=3	-	0.7702	0.7942
true $C_{\text{I},1}$	0.8603	0.8603	-
perceived $\widehat{C}_{\text{I},1}$	0.8643	0.7759	-
true $C_{\text{I},2}$	-	1.0266	-
perceived $\widehat{C}_{\text{I},2}$	-	0.9795	-
true $C_{\text{II},2}$	0.8225	-	0.8225
perceived $\widehat{C}_{\text{II},2}$	0.8261	-	0.6921
true C_{II}|graduate	0.8225	0.9092	1.2257
perceived $\widehat{C}_{\text{II}}$|graduate	0.8261	0.8853	1.1393

Table 1.20: Educational outcomes by school effects

	Low parental education				High parental education			
	p50	p80-p20	p75-p25	p70-p30	p50	p80-p20	p75-p25	p70-p30
repeat level I	0.470	0.273	0.214	0.200	0.069	0.097	0.071	0.052
dropout at t=1	0.113	0.111	0.087	0.048	0.010	0.013	0.010	0.006
drop. at t=2|do not grad.	0.210	0.120	0.091	0.062	0.072	0.035	0.028	0.018
graduate	0.742	0.203	0.140	0.097	0.984	0.025	0.015	0.008
enrol in high school	0.338	0.156	0.137	0.104	0.908	0.067	0.050	0.038
enrol in h.s.|graduate	0.469	0.207	0.143	0.108	0.927	0.059	0.042	0.030
$C_{\text{I},1}$	-0.398	0.464	0.359	0.308	0.646	0.558	0.431	0.370
C_{II}|graduate	-0.300	0.379	0.293	0.234	0.646	0.398	0.316	0.272

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

Table 1.21: Educational outcomes by peers at school

	Low parental education				High parental education			
	p50	p80-p20	p75-p25	p70-p30	p50	p80-p20	p75-p25	p70-p30
repeat level I	0.461	0.376	0.336	0.254	0.067	0.112	0.102	0.079
dropout at t=1	0.125	0.107	0.089	0.060	0.010	0.019	0.012	0.009
drop. at t=2|do not grad.	0.218	0.099	0.080	0.058	0.075	0.037	0.027	0.020
graduate	0.739	0.200	0.175	0.154	0.984	0.030	0.022	0.016
enrol in high school	0.330	0.226	0.195	0.133	0.905	0.085	0.075	0.042
enrol in h.s.|graduate	0.444	0.171	0.141	0.109	0.920	0.069	0.049	0.033
$C_{\mathrm{I},1}$	-0.383	0.332	0.295	0.273	0.665	0.332	0.295	0.273
C_{II}|graduate	-0.315	0.230	0.194	0.182	0.641	0.218	0.191	0.177

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

Table 1.22: Simulated outcomes for the entire sample. School effects raised at the median.

	Baseline	Counterfactuals				
school effects modified		$\mathcal{A}_{\mathrm{I}}, \mathcal{A}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}$	$\mathcal{T}_{\mathrm{II}}$	$\mathcal{T}_{\mathrm{I}}, \mathcal{T}_{\mathrm{II}}$	$\mathcal{R}_{\mathrm{I}}$
repeat level I	0.230	0.215	0.230	0.230	0.230	0.210
dropout at t=1	0.083	0.072	0.071	0.080	0.057	0.081
grad at t=2| in II	0.921	0.927	0.919	0.920	0.917	0.918
drop. at t=2|not grad.	0.212	0.200	0.194	0.203	0.165	0.214
grad at t=3| in II	0.850	0.857	0.841	0.849	0.832	0.847
graduate	0.839	0.857	0.850	0.843	0.865	0.844
enrol in high school	0.659	0.683	0.664	0.683	0.711	0.665
enrol in hs|graduation	0.785	0.797	0.782	0.810	0.822	0.787
$C_{\mathrm{I},1}$	-0.000	0.099	-0.000	-0.000	-0.000	-0.000
C_{II}|graduation	0.272	0.313	0.264	0.269	0.250	0.269

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

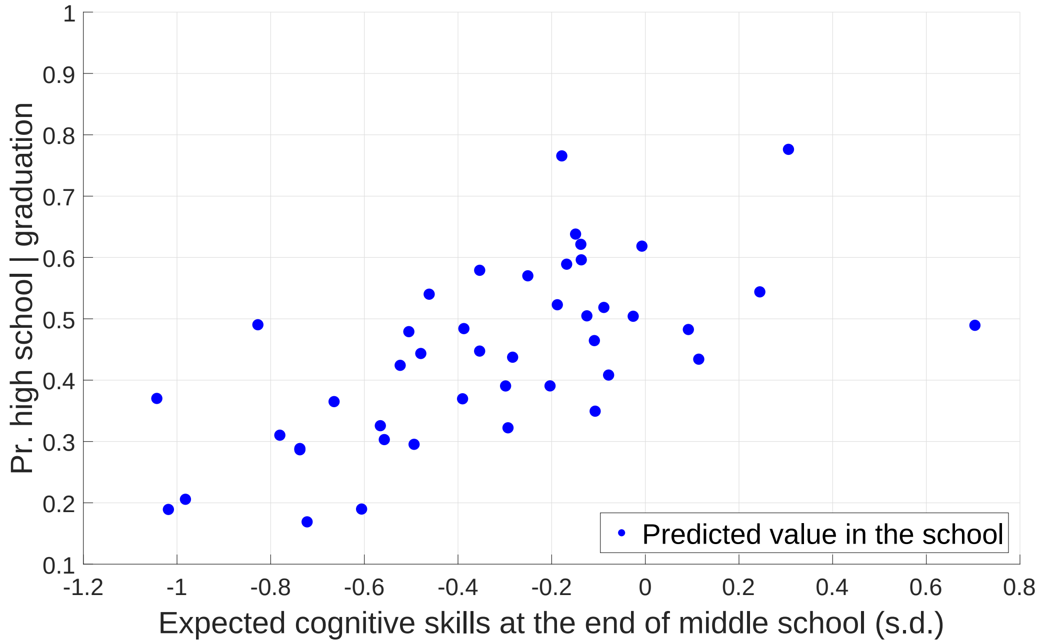
Table 1.23: Simulated outcomes on the subsample of students with low educated parents.

	Baseline	Counterfactuals				
school effects modified		$\mathcal{A}_{\rm I}, \mathcal{A}_{\rm II}$	$\mathcal{T}_{\rm I}$	$\mathcal{T}_{\rm II}$	$\mathcal{T}_{\rm I}, \mathcal{T}_{\rm II}$	$\mathcal{R}_{\rm I}$
repeat level I	0.427	0.385	0.427	0.427	0.427	0.370
dropout at t=1	0.165	0.130	0.131	0.155	0.124	0.159
grad at t=2| in II	0.850	0.878	0.844	0.848	0.843	0.838
drop. at t=2|not grad.	0.251	0.225	0.221	0.235	0.207	0.252
grad at t=3| in II	0.815	0.840	0.799	0.814	0.799	0.811
graduate	0.683	0.742	0.711	0.695	0.721	0.699
enrol in high school	0.399	0.467	0.409	0.455	0.467	0.412
enrol in hs|graduation	0.584	0.629	0.575	0.654	0.648	0.590
$C_{\rm I,1}$	-0.816	-0.628	-0.816	-0.816	-0.816	-0.816
$C_{\rm II}$|graduation	-0.352	-0.242	-0.365	-0.357	-0.370	-0.356

Note. Average outcomes in the column "baseline" are computed using the estimated parameters of the model. In the counterfactuals, the school effects reported above each column are replaced with values at the 75 percentile if they are lower. This table exploits results of the simulations presented in Table 1.13, but frequencies are computed using only the subsample of students with low educated parents (both mother and father have at most lower secondary education).

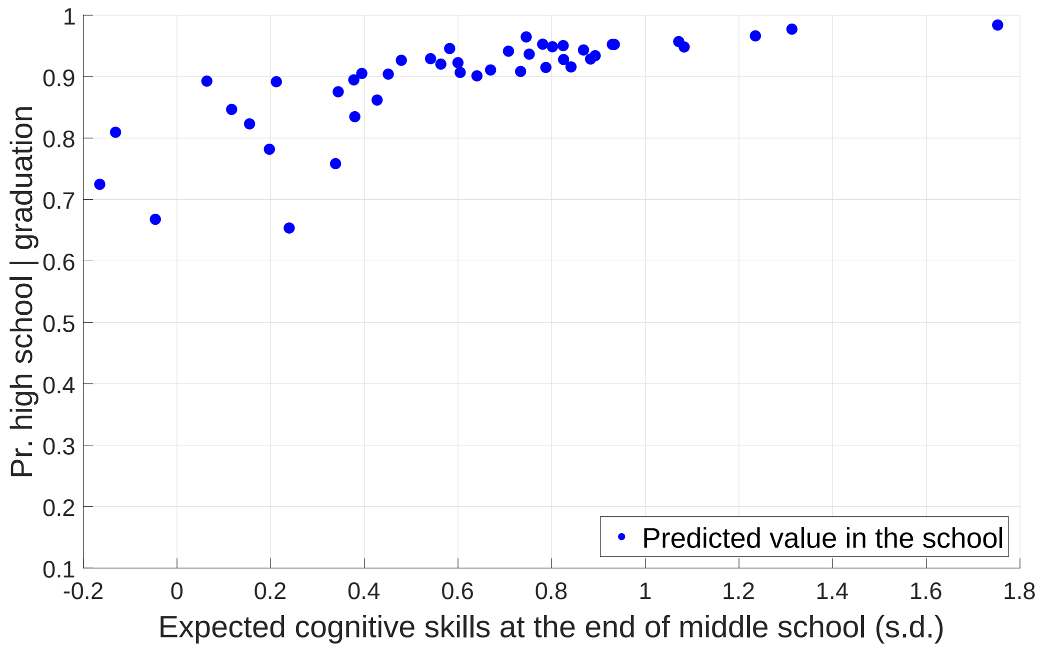
Appendix 1.B Additional figures

Figure 1.8: Student with low educated parents: expected outcomes by school



Note. The figure plots simulated outcomes for a Spanish male student of average innate ability whose parents are low educated (at most lower secondary education). Expected outcomes at the school level are computed using data of the simulation described in Section 1.6.2 and Table 1.11.

Figure 1.9: Student with highly educated parents: expected outcomes by school



Note. The figure plots simulated outcomes for a Spanish male student of average innate ability whose parents have tertiary education. Expected outcomes at the school level are computed using data of the simulation described in Section 1.6.2 and Table 1.11.

Appendix 1.C EM algorithm: theoretical framework

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

$$\log L(o_i;\zeta,\sigma) = \log \int L(o_i;\zeta,\sigma|h)\boldsymbol{\phi}(h)\,dh \tag{1.53}$$

$$L(o_i;\zeta,\sigma|h) = L(r_{i,0};\zeta,\sigma|h)L(g_{i,1};\zeta,\sigma|h)...L(o_{i,T_d};\zeta,\sigma|h) \tag{1.54}$$

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

$$L(r_{i0};\zeta,\sigma|h) = \frac{1}{\sqrt{2\pi\rho_0^r}}\exp\left(-\frac{(r_{i,0}-h-z'_{i,0}\beta_0)^2}{2\rho_0^r}\right) \tag{1.55}$$

Taking the log of (1.54) would simplify the expression and allow an easy estimation through maximum likelihood. Unfortunately the integral over h prevent us from doing so. The proposed approach aims at overcoming this issue.

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

$$\frac{\partial}{\partial\zeta}\sum_i \log L(o_i;\zeta,\sigma) = \sum_i \frac{1}{L(o_i;\zeta,\sigma)}\int \frac{\partial L(o_i;\zeta,\sigma|h)}{\partial\zeta}\boldsymbol{\phi}(h)\,dh = 0 \tag{1.56}$$

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

$$\boldsymbol{\psi}_i(h) = \frac{L(o_i;\zeta,\sigma|h)\boldsymbol{\phi}(h)}{L(o_i;\zeta,\sigma)} \tag{1.57}$$

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 (1.56) can be rewritten as

$$\sum_i \int \frac{L(o_i;\zeta,\sigma|h)\boldsymbol{\phi}(h)}{L(o_i;\zeta,\sigma)}\frac{1}{L(o_i;\zeta,\sigma|h)}\frac{\partial L(o_i;\zeta,\sigma|h)}{\partial\zeta}\,dh = \tag{1.58}$$

$$= \sum_i \int \frac{1}{L(o_i;\zeta,\sigma|h)}\frac{\partial L(o_i;\zeta,\sigma|h)}{\partial\zeta}\boldsymbol{\psi}_i(h)\,dh = \sum_i \int \frac{\partial}{\partial\zeta}\Big(\log L(o_i;\zeta,\sigma|h)\Big)\boldsymbol{\psi}_i(h)\,dh = \tag{1.59}$$

$$= \frac{\partial}{\partial\zeta}\left[\sum_i \int \log L(o_i;\zeta,\sigma|h)\boldsymbol{\psi}_i(h)\,dh\right] = 0 \tag{1.60}$$

Thus if $\widehat{\zeta}$ solves equation (1.56) it solves also equation (2.13) 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

$$\zeta^k = \arg\max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma^k | h) \boldsymbol{\psi}_i^k(h)\, dh \tag{1.61}$$

The general theory ensures convergence of the algorithm.[45]

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

Chapter 2

Grading On A Curve: When Having Good Peers Is Not Good

2.1 Introduction

Student's grades are used for two main purposes: to certify the mastery on a given subject and to compare students when selecting them into tracks, colleges or jobs. We distinguish between tests designed and graded by teachers teaching the subject in school and those tests designed and graded by external examiners, such as centralized authorities nationally or internationally. Tests are usually divided into different questions and each one of them is assigned a number of points. The final grade is then calculated as the percentage points earned out of the total points in the exam. This is clearly the process followed when grading external evaluations, but is less clear-cut when teachers are grading.[1]

Internal evaluations capture human capital accumulation (cognitive skills), as external evaluations do, but may also capture teachers' bias. Lavy (2008), Lavy and Sand (2015) or Rangvid (2015) provide empirical evidence that teachers exhibit a gender bias, often providing differential grades to females or minorities, and show that this bias may have long run effects. Diamond and Persson (2016) uses data from Sweden to show that teachers may inflate grades in high stakes exams for students who had a "bad test day", but do not discriminate on immigrant status or gender. They also show that teacher discretion has long term consequences for individuals in terms of level of education and earnings.[2]

This paper provides empirical evidence of an additional source of disparity between internal and external grades and a channel through which having better peers *need not* be

[1]For instance, an article in The New York Times, "A's for good behavior" notes that teachers often reward students for their good behavior and not for their mastery in the subject. See the full article at http://www.nytimes.com/2010/11/28/weekinreview/28tyre.html

[2]As shown in Betts and Grogger (2003) and Figlio and Lucas (2004) grading standards by teachers may also impact student performance and attainment.

beneficial. In particular we show that a student in a classroom with better peers receives lower grades from the teacher than an identical student with worse peers. In principle in Catalonia – similarly to many other countries – grades in a class do not have to fit a given distribution, but shall measure absolute performance. In practice, the difficulty of lectures and exams may be at least partially adapted to the characteristics of students in the group, and teachers may be induced to grade differently depending on the quality of their students. In this paper we use a minimal definition of *grading on a curve* (GOC). We say that teachers *grade on a curve* whenever having better performing peers harms the grade provided to a given student, namely when relative performances affect the given evaluation.

Providing empirical evidence on these facts presents large challenges, both in terms of identification and data requirements. Using a rich data set of the universe of children in primary and secondary school in public schools in Catalonia we show that grades assigned by teachers are negatively affected by average peer quality.[3] We show that internal scores given by teachers are decreasing in the class average of external evaluations. That is, the internal evaluations are smaller with respect to the external evaluations as peer performance in external evaluations increases. This suggests that teachers value students less in a classroom with better peers. We control for school fixed effects to address selection of students into schools and exploit the fact that students in primary school are homogeneously distributed into classrooms based on time-invariant observables. For secondary school we cannot rule out sorting into classes, but we run a set of robustness checks to confirm the persistence of our results throughout primary and secondary school. In particular, we run a specification with teacher fixed effects and one with school level peer effects, where variation is then captured within a teacher or a school over the years. Our main specification uses as dependent variable the difference between individual internal and external evaluations, and relies on the assumption that they measure the same cognitive skills: we run our analysis on the subset of schools for which the rank generated within class using internal and external evaluations are most similar. We also run the analysis excluding schools with best and worst average external evaluation. The robustness checks confirm our initial findings.

Finally we discuss an alternative specification that allow us to relax the assumption that internal and external evaluations are capturing similarly individual skills. In this alternative specification internal grades are regressed on external grades and average external grades; to address the endogeneity of the variables we propose an instrumental variable approach. Age at enrollment in primary school and its average in the class are

[3]Catalonia is one of the most prosperous autonomous communities in Spain with more than seven million citizens. The Catalan government has the power to legislate in matters such as health or education, among others.

used as instruments for individual external evaluation and mean external evaluations. This specification allows us to directly test that internal and external grades are indeed equally capturing human capital. Results confirm our previous findings.

One of the most widely studied topics in Economics of Education is that of peer effects and how class composition may affect human capital accumulation. The literature is large and the evidence varies – see Sacerdote (2011) and Epple and Romano (2011). But in most studies, having relatively better peers is not harmful on average for human capital accumulation, and is beneficial for most individuals.[4] This paper highlights a potentially different channel through which peer composition can affect long run educational outcomes. Although the accumulation of human capital is not harmed by the presence of better peers, the *perception* that teachers have of students can be affected by the quality of their peers. In particular, if teachers somehow grade on a curve, then having better peers can induce teachers to give lower grades to a given student when faced with better peers.

What are evaluations in school important for? First, they provide feedback to students and their family on their ability and proficiency in a given subject. The informational content of grades and the attitudes of teachers towards individuals in class can affect students' self-image and self-confidence, and substantially influence their future educational outcomes. Such mechanisms have been widely documented in the psychology and sociology literature. Similarly Kinsler, Pavan, and DiSalvo (2014) shows that relative performance of a child in school affects parents' inference of the child's ability and parental investment in the child. Bobba and Frisancho (2016) show that students' perception of their own ability is affected by performance in exams.[5] Azmat and Iriberri (2010) and Tran and Zeckhauser (2012), on the other hand, show that students care and react to their relative position in the classroom; Tincani (2015) shows that rank concerns may generate heterogeneous peer effects; Weinhardt and Murphy (2016) shows that the rank of a student in a class in primary school impacts performance in secondary school, when peers and teachers have changed – they provide survey data consistent with the reason being that students who are high in the rank improve self-confidence and therefore performance in the future. Similarly, Elsner and Isphording (2017) show that rank in the classroom has long run effects on achievement. Hence, grades in the classroom may affect students' perceived ability, future expectations and performance.[6] Our work speaks to

[4]Burke and Sass (2013), Carrell, Sacerdote, and West (2013), and Feld and Zölitz (2016) find that a higher share of top performing peers in the group may harm performances of the low ability students. Our setting is quite different because internal evaluations of every type of students are negatively affected by the presence of better peers.

[5] Ahn, Arcidiacono, Hopson, and Thomas (2016) show that grading policies in college may affect major choice.

[6]Mayer and Jencks (1989) reviews the sociology literature and states that living in an advantageous

this literature providing evidence that some of the signals that students receive on their ability may be distorted by peers' composition and school standards. This may affect their self-perception and educational choices even if their assessment is unrelated to their position in the class.

But evaluations in school can also matter directly to the extent that they determine later access to school track or university. For instance in Germany or Romania, school track in secondary school depends on internal grades. Similarly, access to an excellence program for high school in Madrid, Spain, depends on the internal grades obtained in middle school. On the other hand, university admissions in Spain, Norway or Chile are determined through a centralized procedure for which a mix of internal and external grades determine priority in choosing major and university. In other countries, such as Germany, Sweden or Italy admission to some selective universities or highly demanded majors depends on a score that incorporates among other components internal grades in high school.[7] Finally applications to selective institutions in USA or Canada typically include GPA in high school. Admission committees might be able to weight this information according to the reputation of the sending institution, but most likely they cannot unravel the effect of occasional variations in peers or teachers quality.

To illustrate the implications that these differences between internals and externals may have, we simulate a selection process that selects on the basis of internal grades and compare it to one that selects on the basis of external grades using our data in Catalonia. We find that the 25% top performing students are very different if selected through grades in internal or external evaluations. In particular, more than 30% of those selected through internal grades do not get selected through external grades; vice-versa more than 30% of those selected through external grades do not get selected through internal grades. Of these initial differences, about one third (10 p.p.) is due to differences in the unexplained components of internal and external evaluations. Most of the remaining gap (from 45 to 70%) is due to grading on the curve and school grading policy. Thus differences in grading standards across schools and classes explain a large part of the differences in ranking using internal and external evaluations. Conversely teachers' biases, such as the gender bias, appear to be less relevant in this case.

In Catalonia internal grades impact academic prospects at the end of high school when applying to university, where priority in the desired major in a particular university is given as a function of a compounded grade composed 60% by average GPA (internal grades) in high school (last two years before university) and 40% by a nation wide exam.[8]

---

neighborhood may be disadvantageous, because a given student will rank worse if in an advantageous neighborhood, which may affect his or her expectations.

[7] The organization of education systems in Europe is described in https://webgate.ec.europa.eu/fpfis/mwikis/eurydice. Information about the Chilean system can be found at www.mineduc.cl.

[8] Students can undertake additional field-specific tests to improve their score. This may reduce the

Hence, students at the end of middle school, before starting high school, may be interested in moving to a school with relatively worse peers to increase internal grades towards university admissions. Changing school within the public system is difficult in Catalonia – see Calsamiglia and Güell (2014) for a description of school choice in Catalonia. Moving is slightly more frequent among students that complete a private or semi-private middle school. Among this subsample of movers, 75% move to a school with relatively worse peers than in the previous school.

Estevan et al. (2014) analyze how the Top Ten Percent Law in Texas can generate desegregation in school because relative performance with respect to your peers is what determines access to university. Here we find that a similar effect may impact school choice at the end of middle school: better peers lead to worse internal grades, which in turn affect college admissions. This leads to some students switching schools in search of worse peers.

In the following section we present a simple model describing how external and internal grades are generated. Section 2.3 describes the data. Section 2.4 contains the empirical strategy and the results. Section 2.5 runs simulations on how the top selected students would change if the different sources of disparity between internal and external evaluations were controlled for. Section 2.6 discusses strategic change of school at the end of low secondary education. Section 2.7 concludes.

2.2 A simple model for internal and external evaluations

In this simple illustrative model we assume that individual human capital at a given point in time is a random variable H with expected value $E(H) = 0$. Let $\overline{H}$ be the average human capital in a class, with $E(\overline{H}) = 0$.

External evaluations measure human capital with some noise:

$$\text{ext} = H + \varepsilon_E \tag{2.1}$$

where $cor(\varepsilon_E, H) = 0$.

We assume that internal evaluations capture the same cognitive skills, but may also be affected by biases or grading standards. For simplicity we include just one bias based on gender F ($F = 1$ if student is a female, $F = 0$ if student is a male).[9] Moreover we allow

weight of average GPA in high school to 50%.

[9] In the empirical analysis we test the presence of biases for several observed characteristics; however including more variables here would just complicate the exposition without providing any further insights.

teachers to consider both absolute and relative performance when they assign evaluations.

$$\begin{aligned} \text{int} &= (1-\xi)H + \xi(H - \overline{H}) + \delta F + \varepsilon_I \\ &= H - \xi\overline{H} + \delta F + \varepsilon_I \end{aligned} \tag{2.2}$$

where the error term ε_I is uncorrelated with both H and F. $\xi \in [0,1]$ and $1-\xi$ are weights given to relative and absolute performances respectively. Let's ignore for now the contribution of the bias δF; if $\xi = 0$, i.e. if only *absolute* performance matters, internal evaluations depend only on individual skills H, and would be completely analogous to external evaluations, except for the error component. On the other hand, if $\xi = 1$, the internal evaluation is based only on relative performance, as measured by the distance from the mean in the class. $\xi \in (0,1)$ means that both absolute and relative performance contribute to the final grade. The underlying interpretation would be that teachers adjust evaluations taking into account the average level of the class, either *ex-ante*, adapting the difficulties of lectures and tests, or *ex-post*, comparing students among them when they are assigning final grades. The magnitude of ξ in equation (3.8) tells us the relevance of grading on a curve in the school system under analysis.

The parameter δ captures the additional reward (or punishment) for student gender F. It is important to stress that we do not take a stand on whether $\frac{\partial H}{\partial F} = 0$ or $\frac{\partial H}{\partial F} \gtrless 0$. If for instance $E(H|F=1) > E(H|F=0)$, this would affect in exactly the same way external and internal evaluations. δF only captures any additional difference due to gender that affects only internal evaluations. For example, if females put in more effort in school and therefore learn more contents, this would boost their human capital, increasing similarly both their internal and their external grades. However, if females, as opposed to males, are quiet in class, and teachers award some extra points for good behavior at the end of the year even if their human capital is not larger, δ would capture this. Hence, δ captures any difference between internal and external for females.

Finally, consider a more general specification than the one given in equations (3.1) and (3.8):

$$\text{ext} = \mu_{\text{ext}} H + \varepsilon_E \tag{2.3}$$

$$\text{int} = (1-\xi)\mu_{\text{int}} H + \xi\mu_{\text{int}}(H - \overline{H}) + \delta F + \varepsilon_I \tag{2.4}$$

Here we allow internal and external evaluations to differ in how they capture the human capital. This specification is equivalent to the previous one (up to "rescaling" the human capital) only if $\mu_{\text{ext}} = \mu_{\text{int}}$.

2.2.1 From the model to the empirical estimation

This subsection shortly discusses how to identify the parameters in the data under the assumptions made in the previous paragraphs. Section 2.4 describes in detail our empirical strategy.

The core of our analysis follows from equations (3.1) and (3.8). Human capital is not observed, while the evaluations are. Hence, we can derive H from equation (3.1) and substitue H and $\overline{H}$ into (3.8):

$$\begin{aligned} \text{int} &= \text{ext} - \xi\overline{\text{ext}} + \delta F + \varepsilon_I - \varepsilon_E + \xi\overline{\varepsilon}_E && (2.5) \\ &= \text{ext} - \xi\overline{\text{ext}} + \delta F + \varepsilon && (2.6) \end{aligned}$$

An obvious issue for the empirical estimation of equation (3.9) is that $\text{cor}(\text{ext}, \varepsilon) \neq 0$, because $\text{cor}(\text{ext}, -\varepsilon_E) < 0$; in other words measurement error would downward bias the ols estimate for the coefficient of ext. Therefore we subtract ext from both sides of equation (3.9), and we use the difference between internal and external evaluations as dependent variable:[10]

$$\text{int} - \text{ext} = -\xi\overline{\text{ext}} + \delta F + \varepsilon \qquad (2.7)$$

We can now consider the correlation between $\overline{\text{ext}}$ and the error. Given that the coefficient of $\overline{\text{ext}}$ (i.e. $-\xi$) is smaller or equal than 0, a negative bias leads to overestimating the magnitude of ξ, while a positive bias leads to underestimating the magnitude of ξ. The correlation is the sum of two terms of opposite sign: $\text{cor}(\overline{\text{ext}}, \varepsilon) = \text{cor}(\overline{\text{ext}}, \xi\overline{\varepsilon}_E) + \text{cor}(\overline{\text{ext}}, -\varepsilon_E)$. On one hand, the first correlation is positive, but given that $\overline{\varepsilon}_E$ is an average over several observations we can expect its variance to be quite small. On the other hand, $\overline{\text{ext}}$ is correlated with the individual error ε_E because ext is part of the mean. Therefore the second correlation is negative and may lead to an overestimation of ξ. It is important to stress that this second issue would not arise if we used leave-out mean, i.e. average evaluations among peers in the class, as often done in the peers effect literature (Sacerdote (2011)). However in the current framework it appears intuitively more sensible to use mean at the class level: teachers have a unique reference point (the "average performance") and

[10] It is worth stressing that although ext is part of the dependent variable and its mean $\overline{\text{ext}}$ is a regressor, the model does not suffer from the so-called "reflection problem" (Manski (1993)); as opposed to the case in which a variable y is regressed on its mean $\overline{y}$, here the expected value of ext is not a linear function of the expected value of the other regressors. There is no mechanical relation between the dependent variable and the expected value of $\overline{\text{ext}}$. To see this, imagine that $\xi = 0$, i.e. internal evaluations measure absolute performances. If human capitals in the class are correlated, increasing the expected value of average external in the class implies that it is more likely to have a higher human capital at the individual level. Under our assumptions this would affect exactly in the same way the expected internal and the expected external evaluation, leaving their expected difference (i.e. the dependent variable) unchanged.

compare each child with it, rather than changing reference point for every student. The size of the bias depends on how informative external evaluations are of the underlying human capital, i.e. on the ratio $\frac{\text{var}(\varepsilon_E)}{\text{var}(\text{ext})}$. This can be easily seen considering the OLS estimator of a univariate regression of (int − ext) on $\overline{\text{ext}}$:[11]

$$-\widehat{\xi} = \frac{\text{cov}(\overline{\text{ext}}, \text{int} - \text{ext})}{\text{var}(\overline{\text{ext}})} = -\xi + \frac{\xi \text{var}(\varepsilon_E)}{N^2 \text{var}(\overline{\text{ext}})} - \frac{\text{var}(\varepsilon_E)}{N \text{var}(\overline{\text{ext}})} \quad = -\xi - \frac{(1 - \frac{\xi}{N})\text{var}(\varepsilon_E)}{N \text{var}(\overline{\text{ext}})} \tag{2.8}$$

The total bias is negative. Note that $\text{var}(\overline{\text{ext}}) = \text{var}(\overline{H}) + 1/N\text{var}(\varepsilon_E)$. We didn't make any assumptions on $\text{cov}(H_i, H_j)$ for two students i, j, but, as far as $\text{cov}(H_i, H_j) \geq 0$, for instance because of positive spillover within the class, $\text{var}(\overline{H}) \geq 1/N\text{var}(H)$, and $1/N(\text{var}(H) + \text{var}(\varepsilon_E))$ is a lower bound for the denominator in (2.8). Then

$$\left(1 - \frac{\xi}{N}\right) \frac{\text{var}(\varepsilon_E)}{N \text{var}(\overline{\text{ext}})} \leq \left(1 - \frac{\xi}{N}\right) \frac{\text{var}(\varepsilon_E)}{\text{var}(H) + \text{var}(\varepsilon_E)} \tag{2.9}$$

It is evident from (2.9) that the closer ext is to H, the closer is the estimated $\widehat{\xi}$ to the true value ξ. In practice, as detailed in section 2.4, we perform all the empirical analysis using both average at the class level and among peers. The difference among the two specifications is negligible, providing evidence that the small bias due to the correlation of $\overline{\text{ext}}$ with the individual error is not a concern.

The identification of ξ in equation (2.7) relies on the assumption that internal and external evaluations measure the same cognitive skills. If coefficients for human capital are different, as in the more general formulation given by (2.3) and (2.4), equation (3.9) would rather be

$$\text{int} = \frac{\mu_{\text{int}}}{\mu_{\text{ext}}} \text{ext} + \delta F - \xi \frac{\mu_{\text{int}}}{\mu_{\text{ext}}} \overline{\text{ext}} + \varepsilon \tag{2.10}$$

This is equivalent to the model (2.7) in differences only if $\mu_{\text{int}} = \mu_{\text{ext}}$, namely if the initial assumption holds. Otherwise a different strategy should be implemented to estimate from the data the coefficient $\frac{\mu_{\text{int}}}{\mu_{\text{ext}}}$, and then the rate of grading on the curve ξ. In section 2.4.3 we discuss an alternative approach that allows us to directly estimate (2.10): we instrument for individual and average external evaluations in order to address the measurement error issues. This also allows us to directly verify that the coefficient of ext is about 1, namely that internal and external evaluations measure the same cognitive skills. Instrumenting for $\overline{\text{ext}}$ addresses also the issue of the regressor being correlated with the error term we

[11] For simplicity we assume $\delta = 0$ in this illustration.

discussed above.

2.2.2 Measurement error at the class level

So far we assumed that errors ε_E and ε_I in equations (3.1) and (3.8) are idiosyncratic. A possible concern is that measurement error at the class level may cause a positive or negative shift of the distribution of nation-wide test scores within a class, while teachers are able to unravel the common shock in their evaluations. This would cause a spurious negative correlation between average external evaluations $\overline{\text{ext}}$ and individual difference $\text{int} - \text{ext}$. For concreteness, consider the case in which the error in the external evaluations is at the class level:

$$\text{ext} = H + \overline{\varepsilon}_E, \tag{2.11}$$

so that

$$\text{int} - \text{ext} = -\xi\overline{\text{ext}} + \delta F + \varepsilon_I - (1 - \xi)\overline{\varepsilon}_E \tag{2.12}$$

For instance, this might happen if the test asks several questions on a topic which was not covered well in class; all the external evaluations in the class would be negatively affected ($\overline{\varepsilon}_E < 0$), and the expected difference between internal and external would be positive even if $\xi = 0$. Conversely if the teachers accidentally went through an example that was exactly the same as an exam question, a positive measurement error would affect all the external evaluations within the class, and $E(\text{int} - \text{ext}) < 0$.[12] Therefore, a OLS estimate of the coefficient of $\overline{\text{ext}}$ would be negative even if there were no grading on the curve.

The instrumental variable approach proposed in section 2.4.3 allows us to address this concern. In fact, by definition an instrument is correlated with $\overline{\text{ext}}$ but not with $\overline{\varepsilon}_E$, thus when a 2SLS approach is implemented, the estimate of ξ is not affected by the class level measurement error.

2.3 Catalan school system and Data sources

Primary school (*Educació primaria*, EPRI) is the first stage of compulsory education in Catalonia; children begin primary school in September of the year in which they turn 6 years old. About 67% of students attend a public school; 30% of them attend a semi-private school, and the remaining a private school outside of the public school system.[13] Normally primary education takes 6 years, followed by 4 years of middle school (*Educació*

[12] We thanks an anonymous referee who suggested this example.

[13] Semi-private schools (*Concertadas*) are run privately and funded via both public and private sources.

secondaria obligatòria, ESO). After successfully completing lower secondary education, students can enroll in upper secondary education for two more years.

Within school students are allocated to classes; students in a given class spend almost all the school time together, take all the core subjects together and therefore are exposed to the same set of teachers and teaching methods.[14] In our sample mean class size is 22.2 for primary school and 25.1 for middle school (medians are 23 and 26 respectively).

The core of our analysis (sections 2.4 and 2.5) focuses on students enrolled in either the last level of primary school or the last level of middle school. To be more specific, we study students enrolled in sixth grade in public primary schools in Catalonia from school year 2009/2010 to school year 2013/2014, and students enrolled in fourth grade in public middle schools in Catalonia from school year 2011/2012 to school year 2013/2014.[15] In section 2.6 we exploit data of students enrolled in last grade of middle school and first grade of high school, in all types of schools.

We exploit data from different sources that provide us with detailed information on enrollment, school progression, academic outcomes and socio-demographic characteristics of Catalan students. The *Departament d'Ensenyament* (regional ministry of education in Catalonia) provided enrollment records for the schools in the region, from preschool to high school. 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 we have data for about 60% of them in 2009/2010.[16]

Basic information (date of birth, school and class attended) are available for children in all types of schools, but more detailed socio-demographic characteristics (such as gender and nationality, special needs) are collected only for children in public schools. Moreover for children enrolled in public school we observe the internal evaluations that they receive at the end of the year for each subject they have undertaken. These final evaluations are assigned by teachers taking into account the progression of the child and her performance in several tests administered during the year.[17] For each class in a public middle school we also observe the identifier of teachers that taught Maths and Spanish in that class during the year; we do not have however any additional information on teacher characteristics.

[14]Mathematics, Spanish, Catalan, and English, which are the core subjects exploited in this work are always attended together by all students in the class. In middle school a small number of elective subjects may be attended by only a subset of students in the class, but their evaluations are not part of this study.

[15]These levels correspond to ages 11-12 and 15-16 respectively.

[16]Some schools initially report data only for their lower grades, covering the entire pool of students only after two or three years. Therefore more data is available for more recent years.

[17]For primary school only evaluations at the end of second, fourth and sixth grade (i.e. at the end of "low", "medium", and "high" cycle of elementary education) are officially recorded in the centralized database and available to us. An evaluation of the child's progression is performed also at the end of first, third and fifth grade, in fact children can be retained one more year in the same level at any point of primary education.

The *Consell d'Avaluació de Catalunya* (public agency in charge of evaluating the educational system) provided us with the results of standardized tests taken by all the students in the region attending 6th grade of primary school and 4th grade of middle school.[18] Such tests are administered in the spring since 2008/2009 for primary school and since 2011/2012 for middle school. They assess basic competence in Maths, Catalan, Spanish and English and are low stakes. 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. We refer to the results in these tests, the grading of which is blind, as *external evaluations*, in contrast with the final evaluations given by teachers in the school, that we call *internal evaluations*. The four 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, although the school can decide to exempt students with special educational needs and children that have lived in Spain for less than two years. Moreover children that are sick one or both days and do not show up at school are not evaluated. We drop from the sample children labeled as children with special educational needs (less than 4%). We include in the analysis only classes in which results of the four tests are available for more than 80% of the children in primary school and for more than 70% of the children in middle school.[19]

Finally we collect information on the student's family background, more specifically on parental education from the Census (2002) and local register data (*Padró*).[20]

All data sources have been merged and anonymized by the Institut Català d'Estadistica (IDESCAT).

Table 3.1 shows some basic descriptive statistics by school year. Figure 2.1 plots histograms that describe the distribution of internal and external evaluations.

[18] More information on these tests can be found in the following website (in Catalan): http://csda.gencat.cat/ca/arees_d_actuacio/avaluacions-consell/

[19] We chose these two threshold in order to keep approximately 80% of the observations for both levels of school. We replicate the analysis in section 2.4 and 2.5 choosing different thresholds (in particular including all classes with more than 70% of test takers for primary school, that allow us to keep 90% of the observations) and results are basically the same.

[20] When the information can be retrieved from both sources, we impute the highest level of education, presumably the most up-to-date information. In the analysis we use dummies for "parental background" based on the average level of education of parents: "low" if both parents are early school leavers, "high" if at least one parent holds a tertiary education degree and the other parent graduated from high school, "medium" for any other case. For single-parent family we use the level of education of the single parent. We couldn't identify any of the parents for 4.5% of children in our sample; for them we use a dummy for "missing parents" in the analysis. Excluding them from the analysis does not modify the results. To compute average level of parental background in the class, we use for each student an index representing the average level of education of parents. The index takes 5 values, from 0 (both parents are early school leavers) to 4 (both parents hold a tertiary education degree).

2.4 Empirical Analysis

2.4.1 Main specification

Our analysis follows from the following empirical specification:

$$(\text{int} - \text{ext})_i = \overbrace{-\xi\overline{\text{ext}}_{c_i} + \sigma_{s_i}}^{\text{grading standards}} + \overbrace{\delta_F F_i + \mathbf{P}_i\delta_P + \delta_M M_i}^{\text{individual-level biases}} \overbrace{+\overline{\mathbf{X}}_{c_i}\beta}^{\text{class characteristics}} + \overbrace{\tau_i}^{\text{year}} + \varepsilon_i \quad (2.13)$$

where student i attends public school s_i in class c_i, and receives internal evaluations int_i and external evaluations ext_i. The dependent variable is the difference between internal and external evaluations $(\text{int} - \text{ext})_i$. The right hand side of equation (2.13) includes average external evaluations in the class ($\overline{\text{ext}}_{c_i}$), school fixed effects (σ_{s_i}), dummies for gender (F_i), foreign born status (M_i), a vector of dummies for parental education ($\mathbf{P}_i$, level of education is low, medium or high, or missing), the average of those characteristics at the class level (vector $\overline{\mathbf{X}}_{c_i}$), and year fixed effects (τ_i).

We study separately students in 6$^{\text{th}}$ grade of primary school and 4$^{\text{th}}$ grade of middle school. Both int_i and ext_i are computed as average of four subjects: Maths, Catalan, Spanish and English (we call this average "GPA" from now on). We use z-scores for each year. Using GPA rather than running separate analysis by subject is particularly convenient for two reasons. On the one hand teachers may not separately assign their evaluation, but often meet and discuss together the performance of each student. Therefore we cannot exclude that the final score in one of the subjects is somehow affected by the results in other subjects. Hence, the GPA may be the most suitable measure of skills. On the other hand the internal grade for each subject can take at most 11 different values, therefore using the GPA improves the variation of the dependent variable.[21] We also discuss results when analyses are performed by subjects. Main findings are unchanged.

The coefficient of $\overline{\text{ext}}_{c_i}$, the average external evaluations in the class, will allow us to estimate the rate ξ of grading on the curve. School fixed effects, σ_{s_i}, capture the differences in grading across schools that are constant over time and across classes. Schools may have different grading policies depending on the average pupils they face or the requirements or objectives that they may fix for the school, which may be orthogonal to pupils' observables.

[21] In primary school the available evaluations are "Insufficient", "Sufficient", "Good", "Very good", "Excellent". In middle school each of these words correspond to an interval of numeric grades between 0 and 10: students receives both an integer grade from 0 to 10 and the wordy evaluation associated with it. Using the same conversion scheme, we assign to each evaluation the midpoint of its interval (and then we take z-scores); thus "Insufficient" is interpreted as 3, "Sufficient" as 5, "Good" as 6, "Very good" as 7.5 and "Excellent" as 9.5. An alternative approach for primary school would be to just use numbers from 1 to 5. If the analyses discussed in this section are replicated using this second approach results are extremely similar.

Unfortunately we cannot disentangle which part of the school fixed effect is determined by the quality of the students and which part depends upon other factors. Our identification exploits only within school variation: we are estimating the impact of classmates' quality on internal evaluation conditional on attending a given school. Both grading on a curve, as measured by class-level variation, and school fixed effects cause students with similar characteristics and ability to have different internal evaluations. In this paper we will refer to their joint effect as the effect of *grading standards*.

Equation (2.13) includes a few individual characteristics: gender, foreign born status, parental education; their coefficients are different from zero if those characteristics directly affect internal evaluations on top of their contribution to human capital.[22] The equation controls also for their class averages (vector $\overline{\mathbf{X}}_{c_i}$). Including regressors in $\overline{\mathbf{X}}_{c_i}$ serves two purposes. First, controlling for any class-level bias due to class composition. For instance if on average classes with higher share of females are a more quiet, and teachers are more lenient with a class in which misbehavior is infrequent, then the coefficient of the "share of female" regressor would capture this. Second, for simplicity in section 2.2 we modeled relative performance as computed using the true underlying human capital. In practice if teachers' biases are not fully conscious they may interfere with their estimation of the average human capital in the class. For instance if teachers on average somehow overestimate females skills, they may set a higher reference level in classes with more females.

Equation (2.13) includes also year fixed effects (τ_i). Standard errors are clustered at the class level to allow for unobserved correlation of errors of students attending the same class.[23]

As discussed in section 2.2.1, we prefer to use class-level average variables, rather than mean among peers in the class, although this might produce a small error in the estimation of ξ. We replicated all the analysis described in the paper using both mean variables and leave-out mean variables. Given that results are extremely similar, we show and discuss here estimates from specifications with mean variables, which are closer to our theoretical framework.[24]

While individual characteristics are clearly exogenous regressors, given that their values are determined before the child begins compulsory education, their average in the

[22] In this paper we use "immigrant" and "foreign born" as synonyms. The dummy M_i takes value 1 if the child does not have Spanish nationality. Strictly speaking she may be born in Spain from immigrant parents.

[23] We prefer to cluster at the class level because this is the fundamental unit of our analysis. One might worry about remaining correlation of errors across classes of the same school; in practice we estimate the standard errors clustering at the school level and the significance level never change.

[24] Results with leave-out mean variables are available upon request to the authors.

class may be endogenous because students are not randomly matched to their peers. The same issue applies to $\overline{\text{ext}}_{c_i}$. In section 2.4.2 we extensively discuss children allocation within classes of the same school and potential issues related to sorting of students across classes in secondary education; we run several robustness checks that confirm our finding.

Results

Columns (2) of table 2.2 present the results of the estimation of equation (2.13) for primary and middle school. For comparison in columns (1) only average external evaluations and school dummies are used as regressors.
Coefficients of average external evalutions are -0.61 for primary school and -0.57 for middle school, both significant at the 1%. Estimates are very similar, just slightly smaller in magnitude in columns (1) where we do not control for other regressors. Having in mind the simple model described in section 2.2 we can deduce from the coefficients of average external evaluations that the estimated rate of grading on a curve $\widehat{\xi}$ in the Catalan school system is more than 50%.

The specifications highlight that females are favored in internal evaluations both in primary and in middle school: being a female increases internal evaluations by 0.15 and 0.36 standard deviations respectively. Children of more educated parents receive a relatively higher score in primary school (kids of parents with University degree have internal evaluations that are on average 0.15 higher than their external), while differences in middle school are smaller in size. There is no effect associated with being foreign born in primary school, while there is a positive premium in middle school.

To correctly interpret the results for the "biases" associated with being female, or foreign born, or parental education, it is important to recall that in this paper we call "bias" any differential effect that individual characteristics have on internal evaluations on top of their contribution to human capital. Then, for instance, the result that having highly educated parents ensures on average a positive premium on internal evaluations in primary school should not be interpreted as evidence that teachers are actively discriminating children on the basis of their parental background. What we can conclude is just that those children have their internal evaluations increased for reasons not directly related to their cognitive skills. For instance in primary school parents are actively involved in the educational process, highly educated parents might be more keen to "lobby" for their children. Moreover highly educated parents might on average emphasize more the importance of behaving in class, or make sure that children always complete their homework: the good attitude of their children may then be rewarded by teachers over and above their actual skills level.[25]

[25] Recent immigrant may face special difficulties in adapting to the Catalan educational system, especially if they were educated abroad for a long time before. Teachers may compensate them when grading,

Figure 2.2 plots the estimated school fixed effects (y-axis) versus the average external evaluations at the school level (x-axis). The figure emphasizes that school fixed effects can be sizeable, and on average schools where external evaluations are higher appear to set stricter grading policies.

Results by subject are shown in table 2.5 of appendix 2.C.[26] Overall the qualitative results discussed in previous paragraphs are unchanged; all coefficients are significant at 1% and similar in size to the estimate for the GPA. However both for primary and middle school the estimated rate of grading on a curve is slightly higher for Catalan and Spanish, suggesting that the languages leave more room for subjective evaluations of teachers, and therefore to comparison among students.

"Biases" show a similar pattern across subjects, in particular there is a positive premium associated with highly educated parents in primary school, and a positive effect of being female both in primary and middle school (particularly large for Maths). Given that the analysis by subject is highly consistent with the main specification, we will focus on GPA from now on.

2.4.2 Potential threats to identification

Sorting of students across classes

In Catalonia, as in most countries, students are not randomly allocated to schools: school composition typically reflects neighborhood characteristics. Our analysis includes school fixed effects, so that we only exploit variation within school across classes over the time period covered in our sample. Therefore variation of regressors measured at the class level comes both from the fact that typically a given school has more than one class per year and from the fact that the school appears in the sample for more than one year. During the short period under analysis (from 2009 to 2013 for primary school, from 2011 to 2013 for middle school) there wasn't any change in enrollment rules or in the demography of the region that may suggest changes in schools' composition. In fact average characteristics at the school level such as parental education or share of immigrant students are highly correlated over time. Thus time invariant fixed effects should control for sorting across schools.

While school's enrollment in Catalonia is highly regulated and based on well know priority criteria, rules on how students shall be allocated across classes in a given school

this would explain the positive premium for being immigrant in middle school.

[26] The limitations of studying subjects separately have already been discussed in section 2.4.1.

are not formally defined.[27] Apparently in primary school classes are particularly designed to be homogeneous in the observables. For instance a primary school with two classes for first graders in a given year allocates female students more or less evenly in the two classes. Moreover administrators and teachers use information provided by preschool educators and parents to allocate children so that each class receive a fair number of children that showed high or low ability in the previous years. To support the anecdotal evidence, we formally test that there is no sorting in primary school, finding evidence that student's characteristics and the class the student is assigned to are statistically independent. Appendix 2.A describes our methodology and results. Therefore although children are not assigned to classes with a random draw, their allocation is balanced and the variation in peers composition across classes can be considered *as good as random* for statistical purposes. If anything we may be concerned that the variation in class characteristics across classes of the same school in a given year is limited. Luckily we are also exploiting variation over time, and – as detailed in appendix 2.A – a variance decomposition confirms that although some characteristics vary more between schools than within, there is reasonable variation also across classes.

Conversely a number of middle schools may sort students across classes based on their previous grades or on their intention to pursue further academic studies in the future.[28] We have no information on how teachers of a given school are assigned to classes. Thus there are at least two dimensions that may interfere with our analysis: first, allocation to classes may not be random, i.e. characteristics of students in a class are sometimes correlated; second, assignment of teachers to classes may not be random, i.e. characteristics of teachers and students in the class may be correlated.[29] While we know that students with similar ability or similar background might be more likely to be together, we have no reason to believe that the assignment of teachers to classes follows a systematic pattern, although we cannot exclude that sorting of some kind takes place. In the following paragraphs we will discuss how these potential issues may affect our estimates and perform robustness checks. Appendix 2.B contains a more formal illustration of the challenges to identification.

[27] Most primary schools in our sample have either one or two classes (about half and half), only 6% have three or more classes. Secondary schools are typically larger: almost 40% have two classes, 30% have three classes, 16% four or more, and the remaining only one.

[28] We performed the same battery of tests described in appendix 2.A using data from middle school. Although for each year a large number of schools have pretty much homogeneous classes, overall the results do not allow us to exclude sorting.

[29] This would be the case for instance if more experienced teachers are given higher performing classes, or, vice-versa, if the best teachers are assigned to group of students that lack behind. Unfortunately we have no information on the characteristics of teachers that work with students in our sample.

Let us first abstract from the matching of teachers to classes. A recurrent concern in the peer effects literature is that the sorting of students across classes may interfere with the identification of peer effects on the outcome of interest.[30] Sorting of students across classes is problematic if peer group composition is correlated with omitted variables that affect the dependent variable: estimated coefficients of group characteristics would spuriously capture the effect of omitted variables on the dependent variable. This is a major issue in a quite common setting in the literature: a test score is regressed on individual and peers' predetermined characteristics, to estimate the "reduced form" effect that characteristics of the group of peers have on individual outcome. Then the estimated coefficient may capture both the true effect of peers on individual performance and the fact that being with peers of given characteristics affects the probability that the individual is a high performer. In particular, if sorting is based on performance, a more able student is more likely to be enrolled in a class with high performing peers. In turn performance is typically correlated with predetermined characteristics such as parental background, thus a high performer is more likely to be in class with students with high parental background: a positive coefficient for the average parental background of the peer group may just be due to the positive correlation of this regressor with unobserved components of individual human capital.

Our setting is different because the dependent variable is the difference between two types of evaluations: if the model in section 2.4.1 is correctly specified, the coefficients of the regressors measure the differential effect that the regressors have on internal versus external evaluations. The fact that regressors at the class level are correlated with individual human capital would not be problematic, precisely because human capital is not a determinant of the dependent variable. An important assumption is that external and internal evaluations are meant to measure the same skills, but internal grades incorporate comparison with peers and "biases" that are orthogonal to cognitive skills. However in practice we cannot exclude that there are unobserved variables related to human capital that affect differently internal and external evaluations, and are not orthogonal to the sorting across classes. In particular teachers may observe and reward non-cognitive skills such as grit or perseverance; for instance given two children of similar cognitive ability, a teacher may decide to award a higher grade to the one that always shows interest in class and puts in more effort when doing homework.[31] The same variables might also be taken into account when students are sorted across classes, to asses whether they can benefit from a more challenging program or their willingness to attend an academic education

[30] See Ammermueller and Pischke (2009) for a discussion of potential issues related to non-random allocation of students to classes.

[31] These variables may be correlated with the controls that we are including in the regressions, thus some of the "biases" may take care of part of their effect. However we cannot claim that the limited number of predetermined characteristics we are using fully account for non cognitive skills.

afterwards. Thus children with high unobserved non cognitive skills would be more likely to attend a class with high performing peers (as measured by external evaluations), and they would be more likely to receive high internal evaluations. In this case the coefficient of $\overline{\text{ext}}_i$ would be upward biased. The more aligned internal and external evaluations are, the smaller the bias.

However, we can claim that our estimates provide a *lower bound* for the true relevance of "grading on a curve" in the system, the true effect being potentially larger than the one we find. In fact we expect the coefficient of $\overline{\text{ext}}_i$ (i.e. $-\xi$) to be negative. In practice $-\widehat{\xi}$ would also capture a spurious positive effect on internal evaluations of being with high performing peers. Thus the estimated coefficient may be smaller in magnitude than the true value, leading to an underestimation of ξ.

We now discuss how non-random assignment of teachers to classes may cause further biases in our estimates. The issue here is that teachers grade their own students, and they may have different attitudes: some may be generally more lenient, other stricter, above and beyond the fact that they may compare students among them to assign grades. This is problematic for identification if the "type" of teacher is correlated with the "type" of class: in this case the estimation of $-\widehat{\xi}$ would be affected by any differential leniency of teachers assigned to "good" or "bad" classes. For instance, if more lenient teachers are more likely to teach in classes of high performers, then the coefficient of $\overline{\text{ext}}_i$ would be upward biased. The most problematic case for our exercise is the negative bias that would arise if strict teachers were systematically assigned to classes of high achievers. In this case students in "good" classes would be given internal grades that are low relatively to their external grades not because they are compared with their peers, but because they have a different type of teacher than students in "bad" classes. As a consequence the true "grading on a curve" would be smaller than the estimated one. Although there is no reason to believe that this very specific assignment of teachers to classes takes place, *ex ante* we cannot exclude it or any other correlation between teacher and students' characteristics.

In this paper we perform and discuss in parallel analyses for primary school and middle school. Concerns related to sorting apply only to middle school, because we have evidence that in primary school allocation to classes is "as good as random" for our purpose. The fact that results are fully aligned provides evidence that having different grading standards across classes and schools is a recurrent feature of the Catalan educational system.

Moreover we use a twofold strategy to ensure identification when working with middle school data. First, we replicate analysis for middle school using teacher fixed effects rather than school fixed effects. For middle school we can identify Maths and Spanish teachers for each class. If stricter teachers are assigned classes of high performing students, then teacher fixed effects, rather than comparison with peers, would be the reason why internal evaluations are lower than external in "good" classes and vice-versa. We can test this alternative explanation adding teacher fixed effects to our empirical specification. If the coefficient of $\overline{\text{ext}}_i$ spuriously captures the positive correlation between strict teachers and well-performing class, then controlling for teacher fixed effects should take it to zero or reduce it substantially. For this robustness check we work with internal and external grades in Spanish and Mathematics. Results are shown in table 2.6.[32]

A limitation of our data is that we observe all the teachers that taught a given class at some point during the school year, but we cannot distinguish the main teacher from substitutes. Thus some of the teachers may have spent only few days with the class, for instance while the main teacher was on sick leave, and have no role in the evaluation of the students. In columns Spanish (1) and Maths (1) we include dummies for all the teachers in the dataset, and we allow for multiple teachers associated with students in the same class. In columns Spanish (2) and Maths (2) the sample includes only classes for which we retrieved a single Spanish or Maths teacher (about 75% of the sample used in columns (1)). Results are qualitatively similar to the analysis by subjects discussed in the previous section (table 2.5). Estimated coefficients are somewhat smaller in magnitude, although the difference is at most 0.08: while there might be some correlation between classes of high performers and strict teachers, we can surely rule out the concern that grading on the curve is only apparent and spuriously capture the assignment of teachers to class.

In the second robustness check we replicate the analysis using school-level rather than class-level regressors. In other words, we broaden the definition of peers, including all the schoolmates enrolled in the same level, rather than just classmates, when computing average regressors. Average performance at the school level is obviously correlated with average performance at the class level, but is probably a less precise measure of the references group that teachers have in mind when grading children. Moreover this measure varies only over time (five years for primary school, three years for middle school) given that we are controlling for school fixed effects. The advantage is that we can completely abstract from issues due to sorting of both students and teachers.

As shown in columns (2) of table 2.7 results are quite close to the baseline model; if anything the coefficient for middle school is slightly larger in magnitude. Standard errors are larger for middle schools but results are still significant at 1%.[33]

[32] Given that only a small minority of teachers change school over time, we cannot include school fixed effects in the regressions. Thus teacher dummies are capturing both the individual teacher effect and the school effect.

[33] We use only schools in which more than 80% or more than 70% (for primary or middle school respectively) of students undertook external evaluations, to make sure that the average external evaluations is

Internal and external evaluations

Even if the allocation of students across classes is as good as random for our purposes, any misalignment between internal and external evaluations in capturing human capital that is not controlled for in specification (2.13) may be a source of concern. Section 2.4.3 discusses an alternative specification which allow us to show that internal and external evaluations on average capture human capital in a similar fashion. The robustness checks on the baseline specification we discuss here provide evidence that our results are not driven by unobserved difference between internal and external evaluations across the distribution of human capital.

First, we perform the same analysis on the sample of classes whose rank correlation between internal and external grades is high.[34] In our model comparison of students among them may "shift" up or down the internal evaluations depending on class composition, but does not change the relative position of students in the class: if the only differences between internal and external grades were grading standards, then the rank of students within a class would be the same using the two evaluations. In fact incorporating other students performance in the final internal evaluations does not modify the relative position in the class. In practice both teacher's biases for given individual characteristics and random errors may change the ranking. Moreover, and more problematic for our analysis, if internal evaluations take into account (non cognitive) skills that are not measured by external evaluations, the order might change. Hence, for the subsample of classes with high rank correlation, the two evaluations are truly aligned measure of human capital; if our results are spurious they should not remain when analysis are performed on such subsample.

Within class rank correlation among internal and external grades is generally large in primary school: the median value is 0.85 (mean is 0.83), 75% of classes have rank correlation higher than 0.79, and 25% are above 0.89. In middle school rank correlation is not as large as in primary school, but it is still sizable: median value is 0.69 (mean is 0.64), 75% of classes have a value higher than 0.55 and 25% are above 0.79.

In the empirical analysis in table 2.8 we include only classes for which rank correlation of internal and external evaluations is higher than the median. Estimated coefficients are fully aligned with the baseline specification in columns (2) of table 2.2.[35] This is a further

a meaningful measure of students quality. Therefore sample size in first columns of table 2.7 is slightly smaller than in table 2.2.

[34] For each class we compute the Spearman's rank correlation coefficient between internal and external evaluations in the class; this is equal to the Pearson correlation between the rank values of those two variables.

[35] Setting alternative thresholds (e.g. rank correlation larger than the 60 percentile, or rank correlation larger than 0.75) deliver very similar results.

confirmation that previous results are not driven by spurious correlation with unobserved variables.

Second, we exclude from the analysis schools in which, at least once, a class has particularly high (or particularly low) mean external evaluation; this robustness check acknowledges that ceiling effects may kick in at different points of the underlying distribution of human capital for internal and external evaluations. To clarify this point suppose that the test administered externally is “easy” so that most students above a given level of human capital achieve a top score, while teachers are able to discriminate among them when assigning internal evaluations. Suppose further that human capitals in the class are positively correlated (not necessarily because of sorting of students, positive peer effects on human capital may generate such correlation). Then high achievers would be more likely to be in a class with extremely high external evaluations and they would also be more likely to have a negative difference between internal and external evaluations. This spurious negative correlation might cause the estimated coefficient of mean external evaluations to be negative. Running the analysis only on “average quality” schools would mitigate or solve this issue: if the effect of grading on the curve is not real, it should not remain in this specification.

Table 2.9 show results of this robustness check. More specifically, we identify top and bottom 15% of classes for each year.[36] In columns (1) of table 2.9 we exclude all schools with at least one class in the top 15% in a given year. In columns (2) we exclude schools with at least one class in the bottom 15%. In columns (3) we exclude those with a class in one of the two tails. Estimations on these subsamples confirm that previous results are not driven by classes who did particularly well or particularly badly in their external evaluations. Overall estimated coefficients are quite similar to baseline results in table 2.2. They are somewhat smaller in magnitude when middle schools with classes in the bottom tail are removed (0.47 rather than 0.57), while removing top-performing classes do not affect the estimates neither in primary nor in middle school.

As last robustness check we verify that results are not driven by particular subgroups of the population. We estimate the baseline model on subsamples of the population, such as males or females, and students with low, medium, or high predicted external evaluations (on the basis of their predetermined characteristics). Estimates are consistent across groups and always significant, confirming that the comparison with peers affect all types of students.[37]

[36] We replicate the analysis with stricter or looser thresholds finding very similar results.

[37] The estimated effect is somewhat larger for boys and for students with low predicted internal, although differences are not large in magnitude. Results are available upon request.

2.4.3 Alternative specification

The model estimated in the previous section relies on the assumption that external and internal evaluations measure the same cognitive skills. In this section we provide a direct estimate of

$$\text{int}_i = \gamma \text{ext}_i - \xi \overline{\text{ext}}_{c_i} + \sigma_{s_i} + \delta_F F_i + \mathbf{P}_i \delta_P + \delta_M M_i + \overline{\mathbf{X}}_{c_i} \beta + \tau_i + \varepsilon_i \tag{2.14}$$

using an instrumental variable approach. As already discussed in the previous section 2.2, ext_i can be correlated with the error term ε_i. Our identification relies on the use of A, student's age at enrollment in primary school, as instrument for external evaluation. This approach is correct if A affects the human capital accumulation, but does not impact differently external and internal evaluations. In the next paragraphs we provide evidence in support of the validity of age at enrollment as instrument.

In $\overline{\text{ext}}_{c_i}$ several observations are averaged together, reducing measurement errors that come from idiosyncratic shocks; however, as discussed in section 2.2.2, $\overline{\text{ext}}_{c_i}$ might be correlated with class-level noise. Thus we will also estimate a two stage least square specification in which both A and its average $\overline{A}$ in the class are used as instruments for ext and $\overline{\text{ext}}$.

The fact that a unique school cut-off date determines when a child can enter school induces large heterogeneity in the age at which a child enters school and the heterogeneity of ages encountered in classrooms, with the older children being up to 20% older than their youngest peers. Older children are substantially more mature than their younger peers, which leads them to initially perform better. Work by Heckman and coauthors shows that early child development is complementary to later learning – see Cunha, Heckman, and Lochner (2006) for a review. Bedard and Dhuey (2006) use international data to show that this early relative maturity effects propagate through the human capital accumulation process and have long run effects for adults. Several papers look at the effects within a country.[38]

The case of Catalonia deems particularly interesting as children are generally not allowed to postpone or anticipate entrance to primary school: virtually every child begins primary school in September of the year in which he or she turns 6 years old. This enrollment rule is quite sharp and exceptions are extremely rare.[39] We can verify using

[38]Fredriksson and Öckert (2014) for Sweden, Puhani and Weber (2007) for Germany, Schneeweis and Zweimüller (2014) for Austria, Black, Devereux, and Salvanes (2011) for Norway, Crawford, Dearden, and Meghir (2010) for England, McEwan and Shapiro (2008) for Chile, Ponzo and Scoppa (2014) for Italy, and Elder and Lubotsky (2009) for the US

[39]Enrollment in primary school was regulated by Decree 94/1992, issued on April, 28 (in Diari Official de la Generalitat de Catalunya (DOCG), núm. 1593 - 13/05/1992) until school year 2008/2009 and by

enrollment data for first grade of primary school that more than 99% of children are compliers.

Calsamiglia and Loviglio (2019), exploiting the same data sources of this paper, provide robust evidence that age at enrollment is an important determinant of educational outcomes throughout compulsory education. Figure 2.3 is based on a replication of their main results about the effect of maturity at enrollment on evaluations over time. For each school level, we regress evaluations on age at enrollment and other individual characteristics, including year and school fixed effects; the figure plots the estimated beta coefficients.[40] The age effect is highly persistent over time, although decreasing in magnitude: *ceteris paribus* being born at the beginning of January rather than at the end of December increases the GPA by 0.56 standard deviations at the beginning of primary school, and by 0.32 standard deviations at the end of it. The gap is still sizeable in middle school, where it decreases at a slower pace.

The fact that the effect of maturity on school outcomes decreases over time supports the hypothesis that the difference in maturity is a strong negative shock at the beginning of formal education, that persists over time because current human capital is built on past human capital. Younger children have a learning disadvantage at the beginning of primary school: all children in a class are exposed to the same educational methods and contents, but they may have different learning capabilities due to different levels of maturity. Thus younger students create a lower stock of human capital in the earlier stage of their school career. Later on the difference in maturity is likely to fade out: a child born in January and a similar child born in December have probably the same ability to learn new contents when they are 12, therefore if they had the same level of human capital from previous period, they would be able to increase it in the same way for next period. The issue is indeed that on average they *do not* have the same level of human capital from previous period: the initial disadvantage is so large that the negative effects propagate over time and the gap is not closed at the end of lower secondary education.[41]

The finding summarized in figure 2.3 are reassuring that A surely affects human capital, but it is unlikely to have any differential effect on internal and external evaluations at the end of primary school or later. In fact a potential concern is that if teachers are aware of the “disadvantage of being young”, they might want to correct for it when assigning the evaluation; in this case the effect of age on internal evaluations would be smaller than

Decree 181/2008, issued on September, 9 (in DOGC núm. 5216 - 16/9/2008) from the following year

[40]Sample for level 6 or primary school and level 4 of middle school is the same we used for the main analysis of this paper. In other levels we use all the students for which we could find internal evaluations in mathematics, Spanish, Catalan and English in one of the year in the time range from 2009/2010 to 2013/2014.

[41]Note that the empirical results support the hypothesis because if children continued to increase human capital at a lower rate, the estimated effect of A would be increasing rather than decreasing over time.

that on external evaluations. Conversely the estimated effect is extremely similar using internal or external evaluations: coefficients are virtually the same for primary school and the confidence intervals largely overlap for middle school.

We discuss now the plausibility of exclusion restriction for the average age in the class. It would be violated if the average age in the class has a direct effect on internal evaluations above and beyond its contribution to external evaluations. However the above results suggest that A does not have a different effect on internal and external evaluations, thus $\overline{A}$ should affect differently individual internal and external evaluations even if A does not. To be concrete suppose that students born in the first months of the years, having had an easy time at the beginning of their school career, grow up more enthusiastic about school. Teachers might not reward enthusiasm directly, but they might appreciate a class with more enthusiastic students because it makes a better work environment, thus they might inflate the grades of everyone in the class. In this case $\overline{A}$ would have a direct positive effect on internal evaluations and the estimated coefficient of $\overline{\text{ext}}$ would be smaller in size than the true one. For an example with the opposite bias, we may suppose that teachers do not inflate grades for younger students, but they somehow acknowledge a disadvantage of being in a class with younger students and inflate everyone's evaluations, disregarding their month of birth. These illustrations sound quite unrealistic, but as a matter of fact we cannot completely exclude that $\overline{A}$ is correlated with some unobserved characteristic of the class that affect differently internal and external evaluations. We will first estimate (2.14) instrumenting only ext with A, and then instrumenting both ext and $\overline{\text{ext}}$ with A and $\overline{A}$. Finding results that are similar among them and aligned with the baseline specification in section 2.4.1 would provide evidence that neither measurement errors nor class-level bias are a major concern in the current framework.[42]

Table 2.3 presents results of 2SLS regressions. First stage estimations are shown in table 2.11. In columns (1) only individual external evaluations is instrumented with A, while in columns (2) both individual and average external evaluations are instrumented using A and $\overline{A}$.[43] As in the baseline specifications, we cluster standard errors at the class level. In all the specifications the estimated coefficient of ext is quite close to 1, although the difference is statistically significant for some of them.[44] Following model 2.10 in section 2.2, ξ can be backed up from the empirical estimation dividing the coefficient of $\overline{\text{ext}}$ by

[42] In middle school $\overline{A}$ is most likely correlated with individual performances because some schools sort students across classes. The discussion in section 2.4.2 and appendix 2.B apply to $\overline{A}$ as well.

[43] A is equivalent to the day of birth, rescaled in the interval $[0, 1]$, so that the value for a child born on January, 1 is 1, while it is 0 for a child born on December, 31).

[44] For middle schools p-values of a test that the coefficient is 1 are 0.08 and 0.19 for specifications in columns (1) and (2) respectively. P-values for primary school are 0.03 and 0.08.

the coefficient of ext. Thus the estimated rate of grading on a curve for middle school is 0.62 according to specification (1) and 0.51 according to specification (2). Both figures are quite close to the value of 0.57 found in our baseline specification. For primary school $\widehat{\xi}$ is 0.62 (specification (1)) or 0.35 (specification (2)). The estimate in specification (1) is very similar to the value of 0.61 found in our baseline specification; the estimate of specification (2) is somewhat smaller in magnitude, however it also has a much larger confidence interval that contains the previous estimates.[45]

We replicated robustness checks described in section 2.4.2 using specification (2), which is the most conservative of the two. All results are very consistent with the estimates in table 2.3, although in few cases coefficients are less precisely estimated.

Overall results in this section confirm the robustness of the finding obtained with our main specification in table 2.2, section 2.4.1. In next section we use results of our main specification to compute predicted internal evaluations and residuals; we replicated the same analysis using the alternative approach described in this section, finding very similar results.[46]

2.5 The impact of GOC on selection processes: a simulation

To gain a deeper understanding of the implications of the differences between internal and external evaluations we simulate a selection process that selects the top quartile of students according to either their internal or external evaluations. On one hand academic performance in primary and middle school do not directly matter for tertiary education, thus this simple exercise is just illustrative of what the impact of selecting people using either school grades or standardized tests can be. On the other hand this setting is particularly suited to study differences between internal and external evaluations because it is unlikely that parents have strategically selected school to affect internal grades of the children.

For each school year we rank students from the best to the worst according to internal and external GPA; ties are broken at random. The best 25% are "selected" while the remaining students are "excluded".[47]

[45] To compute formal test of the significance of the difference between the estimate of the baseline specification and each of the alternative specifications, we bootstrapped the statistics 1000 times (resampling classes, which are the cluster unit). We cannot reject the null hypothesis that the difference is 0 for specification (2) (both for primary and middle school). Specification (1) is more precisely estimated, and the null hypothesis is rejected at 5% (not at 1%), although the difference is small in size, especially for primary school.

[46] Results obtained with the alternative specification are available upon request. Footnote 51 briefly compares them with the finding described in the main text.

[47] For a limited number of students (less than 1% every year) the random draw can make a difference

In primary school 31% of students selected through internal evaluations do not get selected through external evaluations and vice-versa; this figure is almost the same (32.5%) for middle school. This sizeable gap suggests that the procedure used to select people can make a difference for a relevant part of the population.[48] However this finding alone does not clarify what is the main reason behind the difference in outcomes of the two procedures.

The empirical model estimated in section 2.4 allows us to interpret the difference between internal and external evaluations as the sum of three main components: grading standards, biases, and residual errors. More specifically, given the estimates reported in table 2.2, columns (2), the individual internal evaluations can be rewritten as

$$\mathrm{int}_i \;=\; \widehat{\mathrm{int}}_i + \widehat{\varepsilon}_i \;=\; \mathrm{ext}_i - \widehat{\xi}\overline{\mathrm{ext}}_{c_i} + \widehat{\sigma}_{s_i} + \widehat{\delta}_F F_i + \mathbf{P}_i \widehat{\delta}_P + \widehat{\delta}_M M_i + \overline{\mathbf{X}}_{c_i}\widehat{\beta} + \widehat{\tau}_i + \widehat{\varepsilon}_i \quad (2.15)$$

where $\widehat{\mathrm{int}}$ is the sum of ext and the predicted difference between internal and external evaluations, estimated following the model discussed in section 2.4.1. The residual $\widehat{\varepsilon}_i$ includes all the differences between internal and external evaluations that are not explained by our model. In particular it contains the difference between the random component of internal and external evaluations $\varepsilon_I - \varepsilon_E$, as it is clear from equation (2.5). Even if biases or differences due to grading standards were not relevant, ranking of students using internal and external evaluations would be different due to different random errors of the two exams. Using $\widehat{\mathrm{int}}_i$ to rank students allows us to get rid of differences between internal and external evaluations due to that randomness.[49] The selection of the top quartile of students performed using external evaluations ext_i and predicted internal $\widehat{\mathrm{int}}_i$ differs for 20.6% of the selected students in primary school and 22.25% of the selected students in middle school. Thus removing the unexplained residual closes about one third of the gap, while the remaining two thirds depends on bias and grading standards, as detailed in equation (2.15).

We can now “shut off” the various components of the RHS in equation (2.15), by simply subtracting the estimated coefficients multiplied by the variables of interest. For instance to get rid off the bias for female we sum $-\widehat{\delta}_F F_i$. Then we redo the rankings, and compare outcomes, to gain further understanding of how each dimension contributes to the difference in outcomes of the two selection processes. Table 2.4 summarizes the re-

between being selected in the top quartile using internal evaluations or being excluded. Results are not sensitive to varying the share of selected students or picking thresholds that ensure that the last selected student is strictly better than the first excluded child. Thus we ignore this issue from now on.

[48] All the results we discuss in this section are weighted averages of the outcomes for each year. Yearly results are remarkably similar.

[49] Being more precise, we also get rid of all the unobserved differences not captured by our model. Thus we can regard the estimated differences in ranking between ext_i and $\widehat{\mathrm{int}}_i$ as a lower bound of the true gap if we could remove only the differences in the random errors ε_I and ε_E.

sults. Interestingly the simulation delivers the same message for both primary and middle school: a large part of the difference is due to grading standards (grading on a curve and school fixed effects), while "biases" are less important.

For primary school removing the effect of grading on a curve reduces the gap by more than 7 p.p. (from 20.6% to 13.5%); eliminating school fixed effects reduces the gap by more than 5 p.p. (from 20.6% to 15.5%). Eliminating both of them simultaneously, namely removing the effect of grading standards from internal evaluations, decreases the gap by more than 13 p.p., explaining 63.8% of the differences between the selection with ext_i and the selection with $\widehat{\text{int}}_i$.

The second part of table 2.4 shows that for primary school getting rid of biases alone would have only minor effects on the gap in rankings. In particular removing the positive bias for female in internal evaluations reduces the gap by 2.5 p.p., while removing the effect of parental education *increases* the gap by about 3 p.p.[50]

For middle school removing the effect of grading on a curve reduces the gap by about 7 p.p. (from 22.3% to 15.3%), while switching off school fixed effects has a slightly smaller effect than in primary school (about 4 p.p.). When both are removed the gap decreases by about 9.5 p.p.; thus grading standards explains 42% of the differences between the selection with ext_i and the selection with $\widehat{\text{int}}_i$.

Biases are slightly more relevant for middle school, explaining 15.2% of the gap – with bias for female having the larger effect. Overall grading standards are by far the most important component. Thus we can conclude that in a selection process of top students at the end of either primary or middle school most of the difference between selection using internal or external evaluations would result from grading on a curve and school grading policies.[51]

[50] This last finding results from the fact that children with higher parental background are disproportionally attending schools with high performing peers, where grading standards as captured by school fixed effects are tougher. Therefore on one hand they are favored by the internal compared with external because of having highly educated parents, but on the other hand they are penalized because they are attending a school that awards less generous evaluations. Whether on average they would prefer to be ranked with internal or external evaluations depend on the relative size of the two opposite effects. However if only biases are removed, there is just the negative effect due to school fixed effects, and we find that the difference between internal and external ranking widen.

[51] We replicated the simulation using the alternative specification described in section 2.4.3, with both student's and class' external evaluations instrumented. Results are qualitatively very similar. The total improvement of removing grading standards is slightly larger when the alternative specification is used (60.5% in primary school and 44.4% in middle school). In primary school school fixed effects are more relevant than grading on the curve (removing them reduces the gap of about 9 and 5 p.p. respectively); this appears consistent with the lower estimated value for $\widehat{\xi}$ using the instrumental variable approach.

2.5.1 GOC and inequality

It is important to understand how different selection systems affect minorities and children with disadvantaged background. We do not observe family income in the data, therefore we rely on parents' education and foreign-born status.

Overall ranking based on internal evaluations select more children from disadvantaged background: students admitted only by the ranking based on the internal but rejected by the ranking based on the external are more likely to have less educated parents and more likely to be immigrant. However they are also attending schools in which peers have less educated parents and have low external evaluations. In other words children with less favorable parental background are more likely to attend "low performing" schools, that inflate internal evaluations more. In fact figures 2.4 and 2.5 suggest that the subgroup of children with less educated parents who attends "high performing" schools would prefer a selection based on external evaluations.

For this analysis we classified schools in 2013 in three groups ("low", "medium", and "high") on the basis of their mean performance in external evaluations the previous years. We can then compare the share of selected students among those with a given parental education and school "quality". Each graph in figures 2.4 and 2.5 focuses on a type of parental background and shows the share of students admitted for each school "quality" type if the ranking is performed using external evaluations (blue bar), and internal evaluations (red bar). Moreover the green bar displays results if internal evaluations are "corrected" by removing residuals and grading standards, as explained in the previous section. The evidence is similar for primary and middle school. For both levels comparing the three graphs we immediately see that under any system and school type children with higher educated parents are much more likely to be selected than students with low educated parents. This is a consequence of the strong correlation between children performances and parental background in Catalonia. However the relative differences between internal and external evaluations are pretty similar across the three graphs: the share of students selected with external in 2013 is clearly increasing in school type, while it is almost flat for internal evaluations. Thus internal evaluations select relatively more children in low-type schools and relatively less children in high-type schools, while the difference is small in medium-type schools.

Looking only at the aggregate statistics we may conclude that overall internal evaluations select more children with low parental background, but this evidence seems to be a consequence of the fact that those students are over represented in schools that we classified as low-type.

2.6 Implications for school choice

In Catalonia internal grades matter when applying to university. In fact priority in the desired major in a particular university is given as a function of a compounded grade composed by average GPA (internal grades) in high school (last two years before university) and by a nation wide exam. The weight given to internal grades is between 50% and 60%, depending on the specific tests chosen in the nationwide exam. According to our previous results, having "worse" peers may be beneficial for internal evaluations. It is then important to understand whether students and their family are aware of this fact and strategically select, when possible, worse quality peers to increase their GPA and boost their chance of admission to the preferred bachelor. This outcome may be suboptimal in terms of human capital accumulation, if they select lower quality schools and end up with less cognitive skills than they would have otherwise.

In most cases the same school that provides lower secondary education also offers high school. If students want to switch to another school they have to apply to a centralized system, and the change is possible only if the chosen school has a free slot. Given that many schools are oversubscribed, especially in the public system, changing school may be difficult in practice. Thus students that we actually observe moving in our data may be a strict subset of the set of students that would be interested in changing, if they could.

For students that move we can compare results of external evaluations at the end of lower secondary education in the old school and in the new school. In particular we can compare the average results in the first and in the second school. Moreover we can assess whether they would have improved their ranking within the school if they attended the new school the year before when the standardized test was administered.

In our data 21% of students in semi-private and private schools change school for the last two years of secondary education; about half of them chose another private or semi-private school, while the other half enroll in a public school. 75% of them move to institutions with lower average results, and 74% of them improve their position in the within-school ranking based on the external test: the average improvement in this subgroup is 18 p.p. Results are consistent when focusing only on students that move to public schools or only on students that move to private or semi-private schools. Moreover it is interesting to notice that although below average performers are more likely to move, individuals across the whole distribution are affected.[52]

Only 11% of students in public schools move to another institution for high school.

[52] The incentives to move depend on the relative grades obtained in one school versus the other and the required grade to access the bachelor of interest to the student. Hence, students moving may be doing so to improve their grade from 8 to 8.25 to enter Medicine or from 6.8 to 7 to enter Economics. This is why we should expect the incentives to matter throughout the ability distribution and not only at one particular threshold.

80% of them stay in the public system, while the remaining 20% select a private or semi-private school. While those in this second group on average slightly improve the quality of their new school, as measured by the average in external evaluations, those that stay in the public system on average move to schools where test scores are lower.[53]

For comparison we check what happens at the end of primary school for those children that have direct access to a middle school in the same institution. Almost all the institutions that offer both primary and lower secondary education are private or semi-private, therefore the most appropriate comparison is with the share of movers from non-public middle school. Note that in this case there is no selection in the near future based on internal grades. Movers at the end of primary school are only 6.5% and on average they chose schools with higher performing peers, contrary to what happens at the end of middle school.[54]

Hence, these results should be taken as suggestive evidence that families understand that there may be grading on a curve and reoptimize their choice in a way that does not necessarily lead to optimal human capital accumulation, but to improved bachelor assignment given that selection depends on internal grades.[55]

2.7 Conclusions

A large challenge when designing school admission procedures is not only to understand what the efficient assignment of resources is, but also to collect the relevant information about individuals to determine that assignment. The main problem is that skills are mostly unobservable. Children spend a large number of years in school, with teachers who spend a large number of hours with them. Hence natural candidates to transmit information about these individuals, their learning capabilities and achieved level is teachers. But teachers may have biases, may respond to stereotypes or have particular preferences which make them less reliable to be the only determinants of a measure of individual skills. That is why in a large number of countries access to tracks, colleges or resources in general is determined using an objective and comparable measure, that is, a nationwide exam. Such countries include South Korea, Brasil or China or India.[56] The problem though is that

[53]However the size of the average difference in quality between old and new school is smaller than what found for students initially attending private schools.

[54]Very few students change school during primary or middle school. Their share is smaller than 4% in all levels.

[55]A rigorous analysis on this would involve estimating preferences by parents under a school choice mechanism that does not provide incentives to tell the truth, which is beyond the scope of the present study. See Calsamiglia, Fu, and Güell (2014) for details on the mechanism used and the challenges that such estimation would entail.

[56]See `portal.mec.gov.br` for Brazil; `en.moe.gov.cn` for China; `mhrd.gov.in/higher_education` for India; `www.kice.re.kr/sub/info.do?m=0205&s=english` for South Korea.

that these exams are very high stakes exams that occur in a small amount of time. Future prospects depeding on a realization of a one day exam does not seem ideal. In addition to the problems that measuring skills in one day exam may presents, there a are a number of papers describing how high pressure can lead to worse performance (Ariely et al. (2009)), how there are heterogenous effects in reaction to such large stakes by gender or race (Azmat, Calsamiglia, and Iriberri (2016) or Attali, Neeman, and Schlosser (2011)), how the levels of pollution can affect the performance (Ebenstein, Lavy, and Roth (2016)). Hence, such large dependence on external exams does not seem efficient nor fair. Other countries have a combination of centralized exams and college-specific selection criteria, such as the US or the UK. Countries such as Spain, Israel, Australia, Denmark, Norway or Sweden have a mixed system, where internal and external grades mechanically produce a rank that determines college assignment.[57]

This paper provides empirical evidence on a novel source of distortion that may arise in any system that uses internal grades to compare students across schools. In particular it puts forth a novel form of grading bias that results when having better peers harms the grades assigned by teachers in school. Internal grades result from teachers following students and evaluating them in a more continuous basis, which seems to be a better evaluating procedure.

On the one hand distortions due to individual or peer characteristics seem unjustified, especially when they may affect the future allocation of students into further career paths. In countries like Israel access to college depends on both internal and external grades but deviations between internal and external evaluations are monitored and penalized at the school level. This may allow for the optimal use of information gathered by teachers, but avoiding systematic and unjustified biases.

On the other hand, grading on the curve indirectly induces effects similar to policies such as the Top Ten Percent Law in Texas, by imposing that rank within class counts towards access to college; or such as other Affirmative Action laws implemented worldwide, that favor children of disadvantages background when accessing further studies. The fact that the impact is achieved through a somehow indirect or unconscious channel can have the added benefit of limiting stigmatizing certain individuals. However, the fact that the impact depends on the school grading policies in general is not ideal if assignment to college, academic tracks or jobs are at stake.

What the optimal balance between internal and external examinations may be requires further understanding of the consequences that such distortions may have. For instance,

[57]See https://webgate.ec.europa.eu/fpfis/mwikis/eurydice for European countries; www.aqf.edu.au for Australia. Lavy (2008) describes school system in Israel.

are those accessing college because of the bias in internal grades benefiting from increased access, are they dropping out more often when in college? Do they benefit more from college than those who were kept out would have? More generally, what are the consequences of these distortions? Similar challenges are faced by the research studying the misallocation costs of Affirmative Action policies and shall be explored in future research.

2.8 Tables

Table 2.1: Descriptive statistics

School year	Parents' education				Female	Immigrant	N students	N schools
	low	*middle*	*high*	*missing*				
primary school – 6th grade								
2009/2010	34.1%	35.8%	25.3%	4.8%	49.5%	12.6%	10,982	372
2010/2011	32.9%	36.4%	26.5%	4.2%	49.7%	12.9%	22,273	764
2011/2012	32.2%	36.6%	27.4%	3.8%	50.0%	11.8%	28,770	927
2012/2013	32.9%	36.6%	26.6%	3.9%	49.5%	12.7%	31,975	1027
2013/2014	31.4%	36.7%	28.0%	4.0%	49.1%	11.6%	33,082	1042
middle school – 4th grade								
2011/2012	35.4%	37.1%	23.6%	3.9%	50.9%	12.6%	22,533	447
2012/2013	35.3%	36.4%	24.1%	4.2%	50.8%	13.1%	25,649	470
2013/2014	35.8%	36.2%	23.9%	4.1%	50.9%	13.9%	25,717	481

Table 2.2: Results

	primary school		middle school	
	(1)	(2)	(1)	(2)
avg external ev.	-0.5789	-0.6097	-0.5352	-0.5662
	(0.0139)**	(0.0140)**	(0.0087)**	(0.0118)**
female		0.1562		0.3605
		(0.0034)**		(0.0060)**
immigrant		-0.0058		0.1986
		(0.0057)		(0.0103)**
parents M		0.0581		-0.0215
		(0.0042)**		(0.0068)**
parents H		0.1522		0.0616
		(0.0047)**		(0.0079)**
Additional regressors	no	yes	no	yes
School fixed effects	yes	yes	yes	yes
N	127,082	127,082	73,899	73,899

Note. Dependent variable is difference between internal and external evaluations at the end of primary school and middle school respectively ("GPA", i.e. average of scores in Mathematics, Spanish, Catalan, English). Sample for primary school spans from 2009 to 2013; sample for middle school spans from 2011 to 2013. Regressors shown in the table are average external GPA in the class ("avg external ev."), and individual characteristics (gender, immigrant, control for parental education (high, middle or low) – parents L, i.e. low educated, is the baseline category). Regressions in columns (2) include also a dummy for observations without info on parental background, average characteristics of the class (share female, share immigrant, average parental education), and year dummies. All regressions include school fixed effects.

Standard errors are clustered at the class level.

Table 2.3: 2SLS regressions

	primary school		middle school	
	(1)	(2)	(1)	(2)
external ev.	1.0384	1.0296	1.1746	1.1217
	(0.0178)**	(0.0169)**	(0.0990)**	(0.0937)**
avg external ev.	-0.6476	-0.3600	-0.7396	-0.5695
	(0.0223)**	(0.1820)*	(0.0986)**	(0.1141)**
female	0.1509	0.1519	0.3570	0.3579
	(0.0041)**	(0.0041)**	(0.0066)**	(0.0064)**
immigrant	0.0061	0.0053	0.2697	0.2490
	(0.0080)	(0.0079)	(0.0416)**	(0.0394)**
parents M	0.0449	0.0460	-0.0559	-0.0482
	(0.0074)**	(0.0073)**	(0.0209)**	(0.0199)*
parents H	0.1268	0.1314	-0.0187	0.0048
	(0.0126)**	(0.0121)**	(0.0461)	(0.0437)
N	127,082	127,082	73,899	73,899

Note. Dependent variable for all specifications is student's internal evaluations (GPA). In columns (1) student's external evaluations is instrumented with student's entry age; in columns (2) student's external evaluations and average external evaluations are instrumented with student's entry age and average entry age. Other regressors, school and year fixed effects are as in table 2.2. First stage estimates are shown in table 2.11. Standard errors are clustered at the class level.

Table 2.4: Selection of top quartile of students

Primary school

Selection based on:		Diff. with external	Improvement
predicted	$(\widehat{\text{int}})$	20.62%	
w/o GOC	$(\widehat{\text{int}} + \widehat{\xi}\overline{\text{ext}})$	13.48%	34.63%
w/o school FE	$(\widehat{\text{int}} - \widehat{\sigma})$	15.49%	24.88%
w/o school FE & GOC	$(\widehat{\text{int}} + \widehat{\xi}\overline{\text{ext}} - \widehat{\sigma})$	7.46%	63.84%
w/o female bias	$(\widehat{\text{int}} - \widehat{\delta}_F F)$	20.11%	2.49%
w/o parents bias	$(\widehat{\text{int}} - P\widehat{\delta}_P)$	21.24%	-3.01%
w/o all individual bias	$(\widehat{\text{int}} - \widehat{\delta}_F F - P\widehat{\delta}_P - \widehat{\delta}_M M)$	20.61%	0.06%

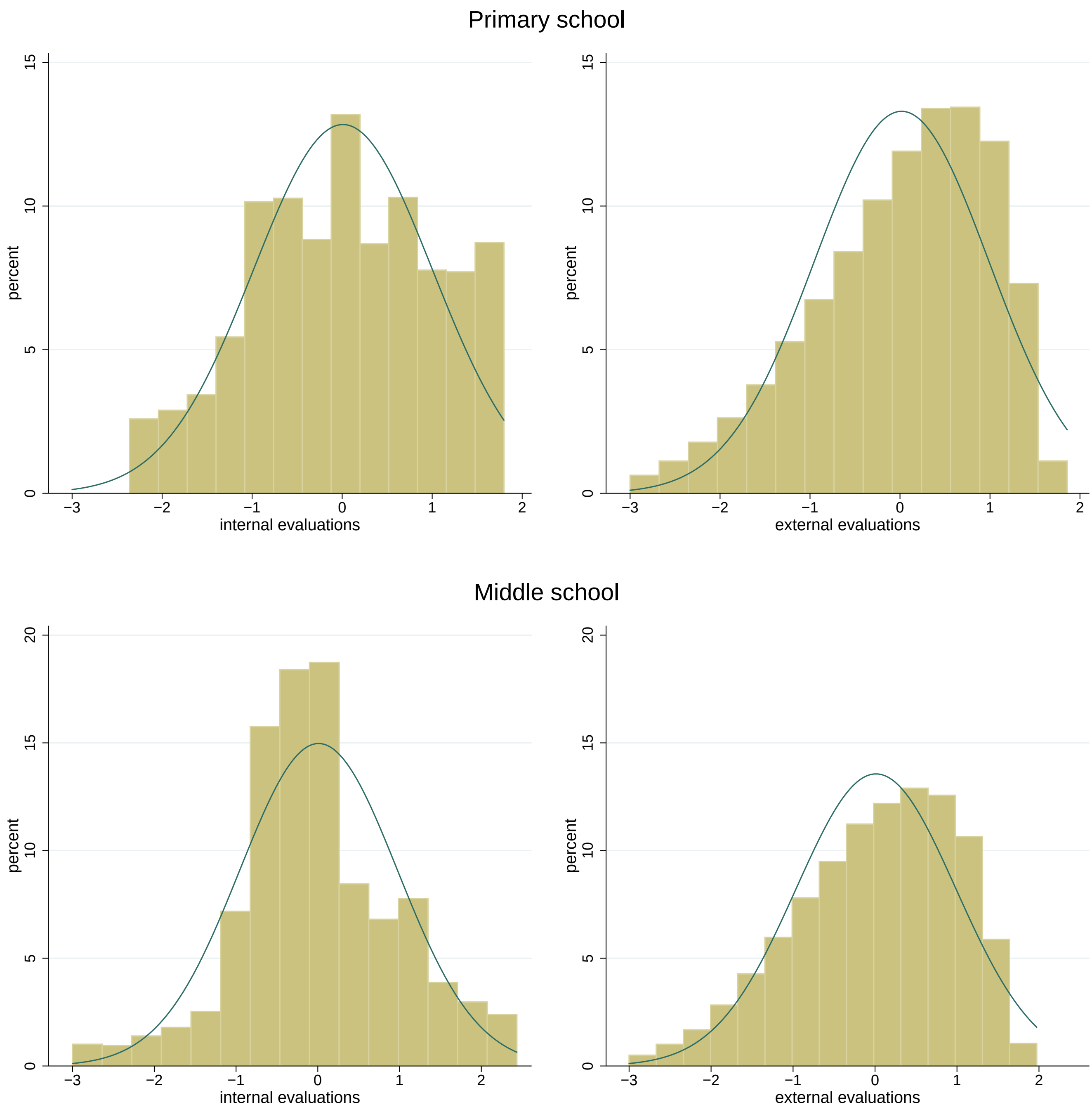
Middle school

Selection based on:		Diff. with external	Improvement
predicted	$(\widehat{\text{int}})$	22.25%	
w/o GOC	$(\widehat{\text{int}} + \widehat{\xi}\overline{\text{ext}})$	15.33%	31.11%
w/o school FE	$(\widehat{\text{int}} - \widehat{\sigma})$	17.99%	19.17%
w/o school FE & GOC	$(\widehat{\text{int}} + \widehat{\xi}\overline{\text{ext}} - \widehat{\sigma})$	12.92%	41.96%
w/o female bias	$(\widehat{\text{int}} - \widehat{\delta}_F F)$	18.97%	14.74%
w/o parents bias	$(\widehat{\text{int}} - P\widehat{\delta}_P)$	22.69%	-1.97%
w/o all individual bias	$(\widehat{\text{int}} - \widehat{\delta}_F F - P\widehat{\delta}_P - \widehat{\delta}_M M)$	18.78%	15.62%

Note. Weighted average of results over time (school years 2009/2010 - 2013/2014 for primary school and school years 2011/2012 - 2013/2014 for middle school). Same samples of table 2.2.

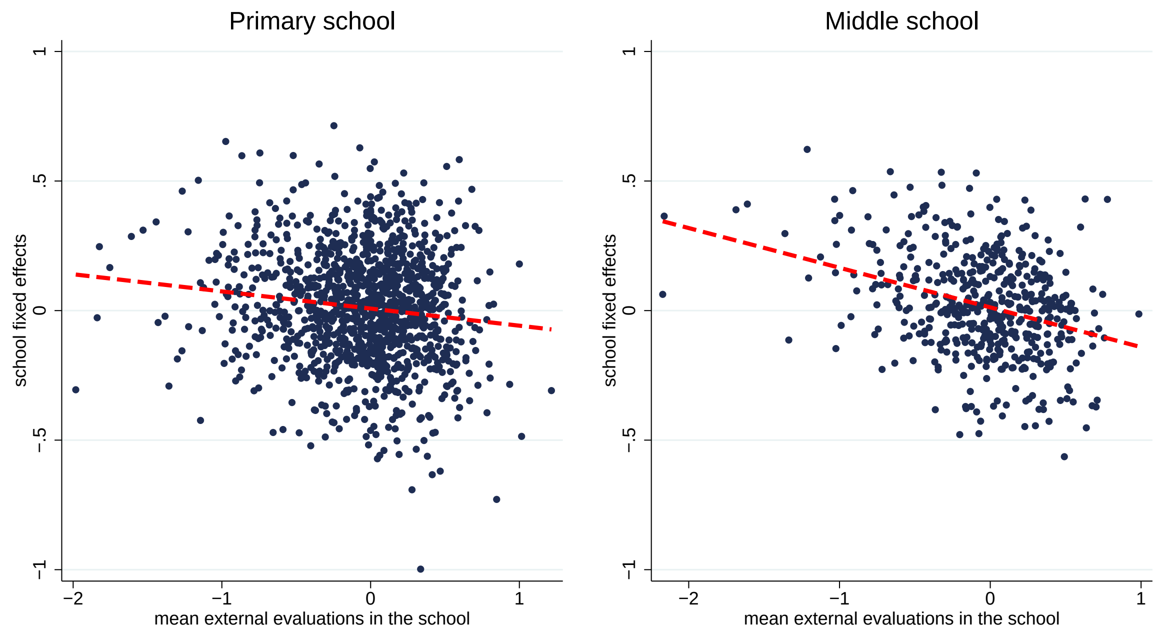
2.9 Figures

Figure 2.1: Distribution of evaluations



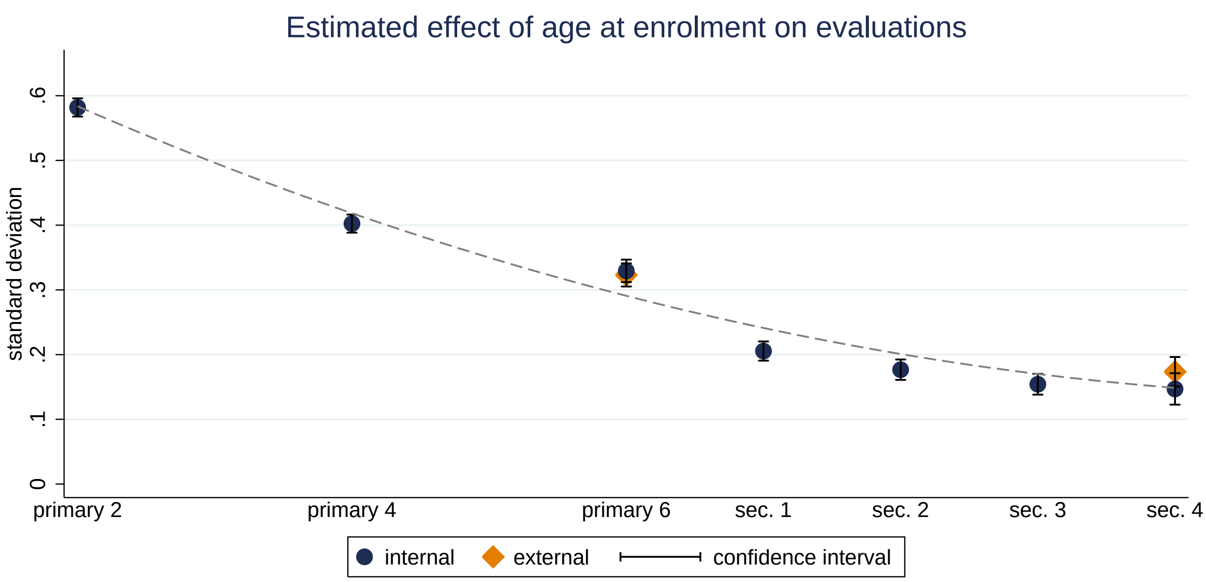
Empirical distribution of internal and external evaluations at the end of primary school (grade 6^{th}) and at the end of middle school (grade 4^{th}). Continuous lines are normal fits. Evaluations are average of four subjects (Maths, Spanish, Catalan, English), standardized at the year level (mean 0, sd 1 among the observations in a given year).

Figure 2.2: Estimated school effects



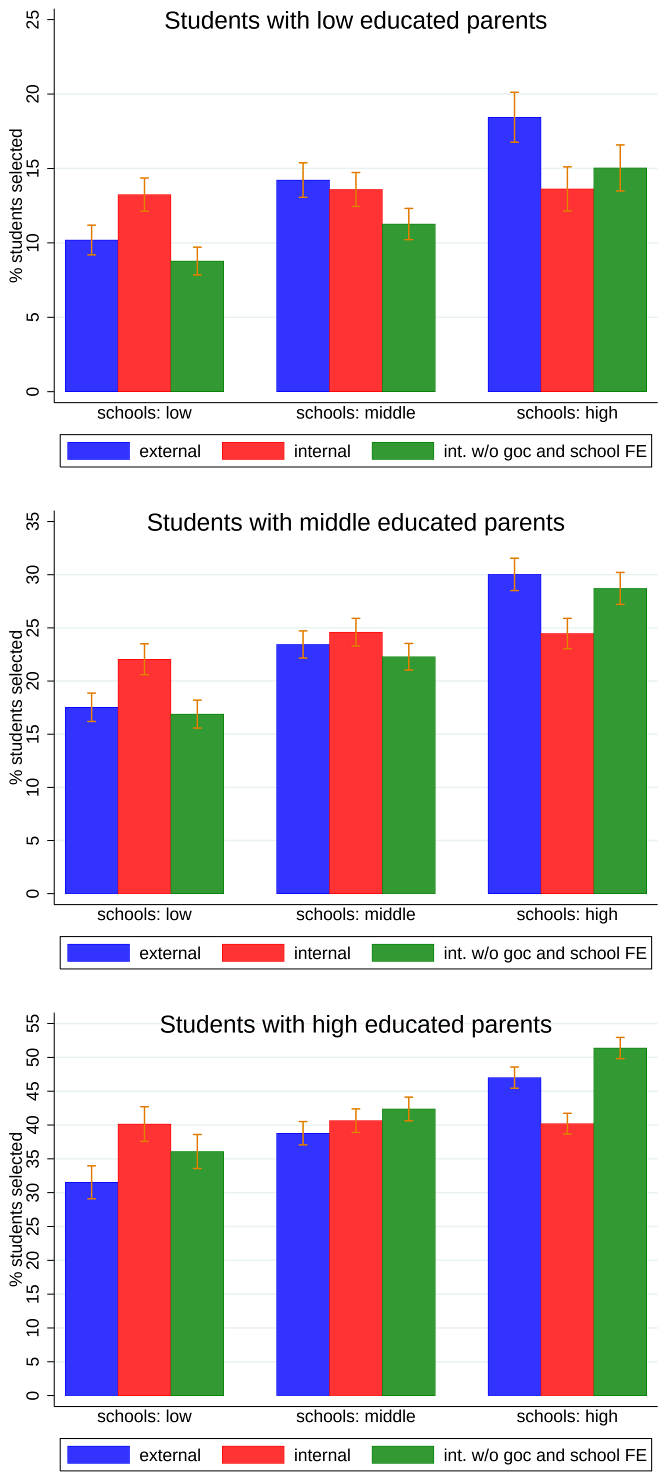
School fixed effects estimated by two stage least square regressions shown in table 2.2, columns (2).

Figure 2.3: Effect of age at enrollment over time



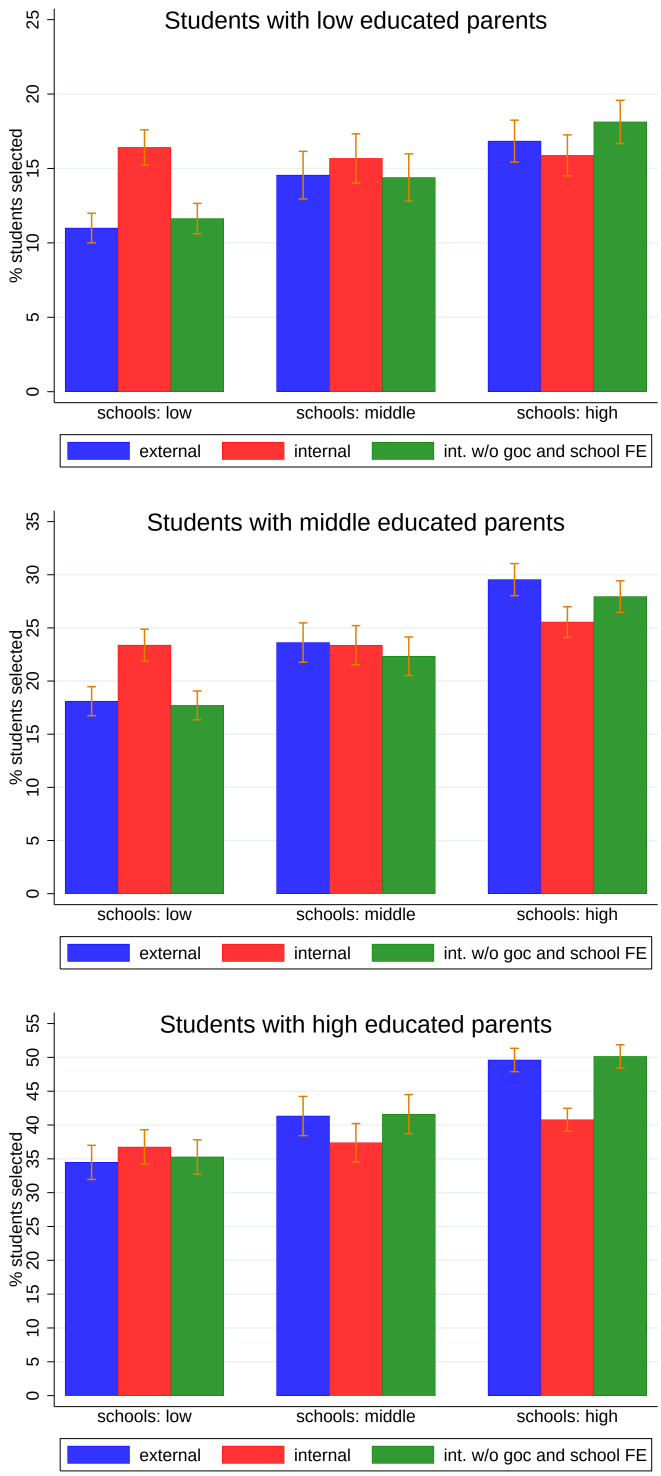
Each dot (diamond) is the estimated coefficient of a regression of internal (external) evaluations on age at enrollment and controls. The dotted line is a quadratic fit of all the estimates.

Figure 2.4: Simulated selection of top quartile of students at the end of primary school



The graphs plot share of admitted students under different selection process in 2013. School quality is defined using school average outcomes in external evaluations from 2009 to 2012: low quality schools are in bottom 33% for at least two years, and never above 66 percentile; high quality schools belong to the best one third for at least two years, and they never belong to the bottom 33%. The top graph concerns students with low educated parents (both attended at most middle school), the bottom graph focuses on students with high educated parents (at least one with tertiary education and the other with high school).

Figure 2.5: Simulated selection of top quartile of students at the end of middle school



The graphs plot share of admitted students under different selection process in 2013. School quality is defined using school average outcomes in external evaluations from 2011 to 2012: low quality schools are in bottom 33% for at least one year, and never above 66 percentile; high quality schools belong to the best one third for at least one year, and they never belong to the bottom 33%. The top graph concerns students with low educated parents (both attended at most middle school), the bottom graph focuses on students with high educated parents (at least one with tertiary education and the other with high school).

Appendix 2.A Class formation in primary school

Although there is no specific regulation on how children should be allocated in primary school, anecdotal evidence suggests that classes are particularly designed to be homogeneous in the observables. For instance a primary school with two classes for first graders in a given year allocates female students more or less evenly in the two classes. Moreover administrators and teachers use information provided by preschool educators and parents to allocate children so that each class receives a fair number of children that showed high or low ability in the previous years. Therefore although children are not assigned to classes with a random draw, their allocation is balanced and the variation in peers composition across classes is *as good as random* for our purposes.

For each primary school in a given year, we can formally verify that there is no sorting, testing whether students characteristics and the class the student is assigned to are statistically independent. Following the procedure described in Ammermueller and Pischke (2009), we perform Pearson χ^2 test for discrete characteristics such as female, immigrant, parental education.[58] Moreover we implement a Kruskall-Wallis test for age at enrollment, which is a continuous variable.

We replicate the same battery of tests for both first and last grades of primary school, to make sure that not only there is no sorting at the beginning of primary school, but that classes are still balanced in sixth grade.[59].

For each characteristic, we reject at 5% level the null hypothesis of "random" allocation less than 4% of times, both in first and in sixth grade. This percentage drops to 0.5% when gender is the characteristic under analysis. We interpret these results as strong evidence that sorting is not in place in primary school; if anything there are interventions to smooth out differences, designing classes to be homogeneous among them.

A natural question that may arise is then whether there is enough variation across classes (and over time) to properly identify the effect of grading on a curve. The variance decomposition in table 2.10 shows that although some characteristics vary more between schools than within, there is a reasonable amount of variation also across classes. The decomposition is computed following Ammermueller and Pischke (2009): first we compute the class averages of each variable, and then we decompose the total variance in these

[58] Given that sample size for each school is relatively small, we also performed Fisher's exact tests, which do not rely on any asymptotic assumption on the distribution of the variables. We find extremely similar results.

[59] In fact some schools shuffle classes either at the beginning of third or of fifth grade. In our sample less than 20% of primary schools do so.

class averages into within school and between school variances.[60]

Appendix 2.B Sorting in middle school

This appendix discusses biases that may affect our estimation when students are sorted across classes. To simplify the notation let us rewrite the model in equation (2.13) as follows

$$\text{int} - \text{ext} = -\xi\overline{\text{ext}} + X\delta + \overline{\mathbf{X}}\beta + \varepsilon \tag{2.16}$$

where we omit the indexes (i and c_i), we ignore school and year fixed effects, and we use the vector X for the individual predetermined characteristics and $\overline{\mathbf{X}}$ for their average at the class level. Our goal is to understand how sorting can bias the estimated coefficients of the class-level regressors, particularly of $\overline{\text{ext}}$.

As explained in sections 2.2 and 2.4, the model in (2.16) relies on the assumptions that internal and external evaluations measure the same cognitive skills, but internal evaluations are modified by comparison with peers and biases that are orthogonal to cognitive skills. However there might be unobserved variables related to human capital (say non-cognitive skills) that may affect differently internal and external evaluations. Moreover we do not explicitly model the fact that some teachers may be generally more lenient or stricter than other, above and beyond the "grading on the curve". The following model incorporates these two aspects in a simple way:

$$\text{int} - \text{ext} = -\xi\overline{\text{ext}} + X\delta + \overline{\mathbf{X}}\beta + \psi N + T_k + \eta \tag{2.17}$$

where N is a measure of non-cognitive skills and T_k is teacher fixed effects, namely a shift up or down of the evaluation depending on the leniency of teacher k. Thus in equation (2.16) $\varepsilon = \psi N + T_k + \eta$. If students are randomly allocated to classes regressor are uncorrelated with ε, and equation (2.16) can be consistently estimated. Conversely if students are sorted across classes, the correlation need not to be 0. Suppose that there are "more difficult" and "easier" classes, and students are allocated based on their cognitive

[60] The formula we use is

$$\frac{1}{N_C}\sum_{s=1}^{S}\sum_{c=1}^{C_s}(x_{cs} - \overline{x})^2 = \frac{1}{N_C}\sum_{s=1}^{S}\sum_{c=1}^{C_s}(x_{cs} - \overline{x}_s)^2 + \frac{1}{N_C}\sum_{s=1}^{S}(\overline{x}_s - \overline{x})^2$$

where x is the variable under analysis, $s = 1, ..., S$ is the school indicator, $c_s = 1, ..., C_s$ is the class indicator (there are C_s classes for school s in our sample), and N_C is the total number of classes in the sample. The first part of the RHS gives the variance within school, the second part the variance between schools. We pool together classes of a given school over time. For instance if school s appears in the sample from 2011 to 2013, with two classes each year, then $C_s = 6$.

and non cognitive skills. Then a student in a class with high average external evaluations probably belong to a "more difficult" class, thus she probably has high cognitive or non cognitive skills (or both). Suppose in class A the external evaluations average to e_A, while in class B their average is e_B, with $e_A > e_B$. Then $E(N|\overline{\text{ext}} = e_A) > E(N|\overline{\text{ext}} = e_B)$. $\text{cor}(\overline{\text{ext}}, \eta) \neq 0$. Abstracting from teacher effects, the correlation is surely positive, and would add a positive bias to the coefficient of $\overline{\text{ext}}$, decreasing its magnitude; thus $\widehat{\xi}$ would underestimate the true effect of GOC.

If students are grouped according to some characteristics, we cannot exclude that the assignment of teachers as well is non-random. This would be a problem if $\overline{\text{ext}}$ (and other average characteristics in the class) are correlated with T_k. In particular if $\text{cor}(\overline{\text{ext}}, T_k) < 0$, for instance because stricter teachers are assigned to high performing classes, then $\text{cor}(\overline{\text{ext}}, \eta)$ may be negative. Consequently the estimated coefficient of $\overline{\text{ext}}$ would be larger in magnitude than the real one, and $\widehat{\xi}$ would overestimate the true effect of GOC.

Appendix 2.C Additional tables

Table 2.5: Estimates by subject

	primary school				middle school			
	Maths	Spanish	Catalan	English	Maths	Spanish	Catalan	English
avg external ev.	-0.653	-0.771	-0.748	-0.670	-0.604	-0.742	-0.718	-0.567
	(0.012)**	(0.012)**	(0.013)**	(0.011)**	(0.014)**	(0.013)**	(0.014)**	(0.012)**
female	0.221	0.161	0.147	0.077	0.489	0.245	0.257	0.191
	(0.004)**	(0.005)**	(0.005)**	(0.004)**	(0.007)**	(0.008)**	(0.007)**	(0.006)**
immigrant	-0.010	0.029	0.017	-0.099	0.163	0.247	0.309	0.031
	(0.007)	(0.008)**	(0.008)*	(0.007)**	(0.011)**	(0.014)**	(0.013)**	(0.010)**
parents M	0.091	0.052	0.051	0.086	-0.004	-0.008	-0.010	-0.008
	(0.006)**	(0.006)**	(0.005)**	(0.005)**	(0.008)	(0.009)	(0.008)	(0.007)
parents H	0.213	0.154	0.148	0.173	0.074	0.108	0.089	0.060
	(0.006)**	(0.006)**	(0.006)**	(0.006)**	(0.009)**	(0.010)**	(0.010)**	(0.008)**
N	127,082	127,082	127,082	127,082	73,899	73,899	73,899	73,899

Note. Dependent variable is difference between internal and external evaluation in the subject reported above the column; "avg external ev." is the average at the class level in the same subject. Other regressors, school and year fixed effects are as in table 2.2. Standard errors are clustered at the class level

Table 2.6: With teachers' dummies

	middle school			
	Spanish (1)	Spanish (2)	Math (1)	Math (2)
avg external ev.	-0.691	-0.698	-0.519	-0.525
	(0.014)**	(0.016)**	(0.015)**	(0.017)**
N	72,056	54,627	72,044	51,259

Note. Dependent variable is difference between internal and external evaluation in either Spanish or Mathematics. Regressions (1) include dummies for Spanish or Maths teachers. Regressions (2) works similarly, but the sample is restricted to classes that have a unique Spanish or Math teacher associated. Other regressors, school and year fixed effects are as in table 2.2. Standard errors are clustered at the class level.

Table 2.7: With school-level mean

	primary school	middle school
avg external ev.	-0.647	-0.662
	(0.014)**	(0.040)**
N	122,951	70,195

Note. Dependent variable is difference between internal and external evaluations (GPA). Average external evaluations is the mean of students' evaluation at the school level in a given year (rather than at the class level). Other regressors and school fixed effects are as in table 2.2 (average are computed at the school level). Standard errors are clustered at the school level.

Table 2.8: Classes with high rank correlation

	primary school	middle school
avg external ev.	-0.616	-0.543
	(0.015)**	(0.020)**
N	63,564	36,960

Note. Dependent variable is difference between internal and external evaluations (GPA). This table shows results of the baseline specification performed on the subset of classes in which the rank correlation between external and internal evaluations is larger than the median (other regressors, school and year fixed effects are as in table 2.2).

Table 2.9: Without outlier schools

	primary school			middle school		
	(1)	(2)	(3)	(1)	(2)	(3)
avg external ev.	-0.6175	-0.5829	-0.5882	-0.5746	-0.4647	-0.5688
	(0.0217)**	(0.0147)**	(0.0284)**	(0.0183)**	(0.0233)**	(0.0368)**
N	70,202	78,029	28,690	34,478	30,598	13,613

Note. Dependent variable is difference between internal and external evaluations (GPA). This table shows results of the baseline specification performed on subsamples of schools. In columns (1) we exclude all schools with at least one class in the top 15% of mean external evaluations in a given year. In columns (2) we exclude schools with at least one class in the bottom 15% in a given year. In columns (3) we exclude those with a class in one of the two tails. Other regressors, school and year fixed effects are as in table 2.2.

Table 2.10: Variance decomposition

Variable	**Between**	**Within**	**Total**
GPA external	0.184	0.076	0.259
	70.8%	29.2%	
GPA internal	0.071	0.056	0.127
	56.2%	43.8%	
A	0.001	0.003	0.004
	18.1%	81.9%	
parents	0.377	0.074	0.450
	83.7%	16.3%	
female	0.002	0.007	0.009
	21.0%	79.0%	
migrant	0.027	0.006	0.033
	81.1%	18.9%	

Table 2.11: First stage estimates

	primary school			middle school		
	(1) ext. ev.	(2) ext. ev.	(2) avg ext. ev.	(1) ext. ev.	(2) ext. ev.	(2) avg ext. ev.
entry age	0.3238	0.3272	0.0002	0.1729	0.1068	0.0046
	(0.0086)**	(0.0087)**	(0.0001)	(0.0112)**	(0.0100)**	(0.0014)**
avg entry age		-0.0830	0.2220		1.4972	1.7221
		(0.0544)	(0.0552)**		(0.1288)**	(0.1395)**
female	0.1385	0.1385	0.0054	0.0524	0.0501	0.0348
	(0.0051)**	(0.0051)**	(0.0012)**	(0.0070)**	(0.0070)**	(0.0040)**
immigrant	-0.3259	-0.3259	-0.0223	-0.5692	-0.5630	-0.1870
	(0.0096)**	(0.0096)**	(0.0024)**	(0.0134)**	(0.0133)**	(0.0094)**
parents M	0.3258	0.3258	0.0177	0.2938	0.2882	0.1202
	(0.0064)**	(0.0064)**	(0.0016)**	(0.0084)**	(0.0083)**	(0.0055)**
parents H	0.6416	0.6415	0.0308	0.6254	0.6192	0.2068
	(0.0071)**	(0.0071)**	(0.0019)**	(0.0104)**	(0.0103)**	(0.0073)**
N	127,082	127,082	127,082	73,899	73,899	73,899

Note. First stages of 2sls regressions shown in table 2.3. “entry age” is the student’s (expected) age at enrollment in first grade of primary school. This variable has been scaled in the interval $[0, 1]$ (it is 1 for a child born on January, 1; it is 0 for a child born on December, 31). “avg entry age” is the mean value at the class level. Other regressors, school and year fixed effects are as described in table 2.2. Standard errors are clustered at the class level.

Chapter 3

Maturity and School Outcomes in an Inflexible System: Evidence from Catalonia

3.1 Introduction

The fact that a unique school cut-off date determines when a child can enter school induces large heterogeneity in the age at which a child enters school and large heterogeneity of ages encountered in classrooms, with the older children being up to 20% older than their youngest peers. Older children will be substantially more mature than their younger peers, which will lead them to initially perform better. Work by Heckman and coauthors shows that early child development is complementary to later learning – see Cunha, Heckman, and Lochner (2006) for a review. Bedard and Dhuey (2006) use international data to show that this early relative maturity effects propagate through the human capital accumulation process and have long run effects for adults. Several papers look at the effects within a country: Fredriksson and Öckert (2014) for Sweden, Puhani and Weber (2007) for Germany, Schneeweis and Zweimüller (2014) for Austria, Black, Devereux, and Salvanes (2011) for Norway, Crawford, Dearden, and Meghir (2010) for England, McEwan and Shapiro (2008) for Chile, Ponzo and Scoppa (2014) for Italy, and Elder and Lubotsky (2009) for the US.

We use very detailed administrative data of the universe of public schools in Catalonia for children between age 6 and 18 for the academic years 2009/2010 - 2013/2014 to provide evidence that being younger affect educational outcomes in several ways. The case of Catalonia deems particularly interesting as children are generally not allowed to postpone entrance to primary school. We find that relatively younger children perform worse as measured by their performance in tests administered both at the school level and at the regional level; moreover they are retained more often and are less likely to enroll in the academic track for high school. Contrary to what others have found for other countries, younger children in Catalonia drop out more often.[1] We also find that the effect is decreasing as children get older, although significant throughout. In particular the gap between younger and older children is almost 0.6 standard deviations when they are in second grade, and still 0.2 standard deviations in middle school. The effect is significant and sizeable for all ability levels, as confirmed by quantile regressions.

Particularly stark are the results on retention: while a small fraction of students are retained in primary school (about 3.1% during the first two years) those born at the end of the year are more than two times more likely to be retained than the average child. At age 14 they are almost 11 percentage points more likely to be in a lower school-grade than their peers born at the beginning of the year.

Younger children are more frequently diagnosed with learning disabilities such as ADD and ADHCD. The probability of being diagnosed for a child born at the end of the year is three times higher than for an otherwise identical child born at the beginning of the year.

Finally, being younger increases the probability of dropout by about 2 percentage points. When children finish middle school, at age 16, they can choose a vocational or an academic track. About 50% of students overall enroll in the academic track, but the probability of choosing it is 5 p.p lower for the youngest.

In the literature there is also evidence that maturity effects may be larger for children coming from families with high socioeconomics (SES). For instance, Elder and Lubotsky (2009) argue that pre-school experience changes with SES and hence the impact of spending one more year at home or in preschool is particularly benefiting for children coming from high SES. In the US, for instance, the impact of maturity is larger for high SES than for low SES. On the other hand high SES could also help fill the gap during school years, by providing support and additional training to the relatively younger who are facing greater challenges. In the case of Catalonia SES of the parents affects performance directly, by increasing performance in school, but does not interact with maturity effects. This may be because more than 97% of the students enroll in preschool at the age of 3, and a vast majority of them attend childcare before that.[2] This means that the extra

[1] In Catalonia children get their degree at the end of the academic year where most students turn 16, which is the age until which education is compulsory. As we discuss later in the text incentives to finish the degree may be slightly stronger in Catalonia than in other countries.

[2] Guaranteed universal and free access to public education in Spain, and in Catalonia in particular, starts at age 3. The spring of the year that children become three parents are asked to choose school for their children. That will be the same school that children will be in at least until the end of primary school. In our data we can observe transition from preschool to primary school only for children enrolled in public schools. Only 4.5% of children who attended a public preschool switch to a different institution for primary education. These figures compare well with statistics described in Calsamiglia and Güell (2014) or Calsamiglia, Fu, and Güell (2014), for both public and semi-public school in Barcelona, where

year out of school for older children is spent in homogeneous preschools for all children, and hence, there is less heterogeneity due to the experience acquired in that additional year.

The rest of the paper is organized as follows. Section 2 describes the Catalan educational system and enrollment rules. Section 3 presents the data and the descriptive statistics. Section 4 presents the analysis of the maturity effects on school outcomes in primary and middle school. Finally, Section 5 concludes.

3.2 Catalan educational system and enrollment rules

Primary education (*Educació primaria*, corresponding to ISCED level 10) is the first stage of compulsory education in Catalonia; normally a child begins primary school in September of the calendar year in which he or she turns 6 years old. In other words, the cutoff date is January 1. For instance, both a child born on January 1, 2003 and a child born on December 31, 2003 starts primary school in September 2009, while someone born on January 1, 2004 starts one year later, in September 2010. This enrollment rule is quite sharp and exceptions are extremely rare. Until 2008, they required the approval of the regional ministry of education.[3] A reform of compulsory education approved in September 2008 introduced a slightly more flexible transition from preschool to primary school. The general rule is the same, but exceptions are managed by the school, together with the family instead of the department of Education.[4] However, as shown in next section, also in more recent years the overwhelming majority of children comply with the rule: more than 99% of them enroll in primary school at 6 years old.

Before primary education children can attend preschool (*Educació infantil*) for three years (from September of the year in which they turn 3). Attendance is not compulsory, but in practice almost all families enroll their children.[5] Normally primary education takes 6 years, followed by 4 years of lower secondary education (*Educació secondaria obligatòria*, corresponding to ISCED level 24). Students are legally required to stay in school until they turn 16 or until they graduate. After successfully completing lower secondary education, students can enroll in upper secondary education for two more years. Upper secondary education is either academic (*Batxillerat*, corresponding to ISCED level 34) or vocational (*Grau Mitjà*, corresponding to ISCED level 35). Only students who complete academic

95% of the children attend school at age 3 and less than 5% change school in primary school.

[3]Decree 94/1992, issued on April, 28 (in Diari Official de la Generalitat de Catalunya (DOCG), núm. 1593 - 13/05/1992).

[4]Decret 181/2008, issued on September, 9 (in DOGC núm. 5216 - 16/9/2008), and Order EDU/484/2009, issued on November 2, (in DOGC núm. 5505 - 13/11/2009)

[5]In our data about 97% of children enrolled for the first time in primary school in 2012 or 2013 were enrolled in preschool the year before. Some of the remaining 3% may have been enrolled in a semi-private or private school (only enrollment data for public preschools are available to us).

upper secondary education can aim at enrolling in a 4-year bachelor degree in a University. The rates of completion of secondary education and enrollment in further education is low in Catalonia (and more in general in Spain), in comparison with other European countries. According to Eurostat in years 2011-2013 about 25% of the population aged 18 to 24 in Catalonia were early leavers from education, namely they attained at most lower secondary education and have not been involved in further education or training.[6]

Students who do not obtain sufficiently good evaluations can be retained and spend one more year in the same grade. By law, children can be retained at most once during primary education and at most twice during lower secondary education.[7] Exceptions are allowed only for children with special educational needs and are extremely rare in practice.[8] However, as shown in section 3.3.2, retention is relatively infrequent during primary education and more common during lower secondary education.

3.3 Data

Our analysis focuses on students enrolled in public schools in Catalonia from school year 2009/2010 to school year 2013/2014. Primary education and lower secondary education is typically provided by different schools. We call "primary schools" the providers of the former and "middle schools" the providers of the latter. Each year more than 50.000 children enroll for the first time in the first grade of a public primary school of the region. They are about 65% of the total number of students who enter primary education in that year.[9]

3.3.1 Data sources and sample selection

We exploit data from different sources that provide us with detailed information on enrollment, school progression, academic outcomes and socio-demographic characteristics of Catalan students. The Institut Català d'Estadistica (IDESCAT) anonymized the data sources and provided us with student identifiers to merge them.

[6]Statistics "Early leavers from education and training by sex and NUTS 2 regions" (edat_lfse_16) available on ec.europa.eu/eurostat

[7]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)

[8]Our data confirm that retention rules are enforced. For instance among children born in 2003 (who normally enter primary education in 2009 and are in grade 5 by year 2013), only 30 (0.05%) were retained twice during the first two cycles of primary education. Among those born in 2001 (who typically complete the second and the third cycle of primary school before 2013), only 27 (0.04%) were retained twice in the time period spanned by our sample. Similarly for middle school, only 13 children born in 1997 and 22 children born in 1996 were retained three times from 2009 to 2013.

[9]More than 30% of the students attend semi-private schools (*Concertadas*), which are run privately and funded via both public and private sources. The remaining students attend a fully private school. For simplicity in this paper we will define as "private" all schools that are not fully public.

The *Departament d'Ensenyament* (regional ministry of education in Catalonia) provided enrollment records for the schools in the region, from preschool to high school. The IT infrastructure that supports the automatic collection of data have been progressively introduced since the school year 2009/2010. By year 2010/2011 almost all schools have already adopted it, while we have data for about 60% of them in 2009/2010.[10] For all students in the public system we observe the school and the class in which she is enrolled, the day of birth, the gender, the nationality, and whether she has special educational needs. Moreover, data include evaluations that the students receive at the end of grades 2, 4, and 6 (primary education), and at the end of each grade from 7 to 10 (secondary education), for each subject that they have undertaken. These evaluations are assigned by teachers taking into account the progression of the child and her performance in several tests administered during the cycle. For students enrolled in grade 10, i.e. the last grade of middle school, we can observe whether they graduated at the end of the year.[11] Data also include some enrollment information and date of birth of students attending a non-public school, but neither other demographics nor their evaluations. Therefore this paper focuses on students enrolled in public schools. Appendix 3.B replicates some of the analyses for students enrolled in non-public schools.

The *Consell d'Avaluació de Catalunya* (public agency in charge of evaluating the educational system) provided us with the results of standardized tests taken by all the students in the region attending the last year of primary school (grade 6) and the last year of middle school (grade 11). Such tests are administered during spring since 2009/2010 for primary school and since 2011/2013 for middle school. They assess basic competences in Maths, Catalan, Spanish and English and have a purely statistical purpose: they do not affect the students' final evaluations or progress to the next grades. We will refer to the results in these exams, whose grading is blind, as *external evaluations*, in contrast with the final evaluations given by the school, that we will call *internal evaluations*.

Finally we collect information on the student's family background, more specifically on parental education from the Census (2002) and local register data (*Padró*). When the information could be retrieved from both sources, we imputed the highest level of education, presumably the most up-to-date information. We could not retrieve information about their parents for 4.5% of children, those students are excluded from the analyses.

[10]Some schools initially report data only for their lower grades, covering the entire pool of students only after two or three years. Therefore the sample used to study the outcomes in the lower grades is slightly larger.

[11]We could not retrieve some or all the evaluations for about 2% of the students, who are excluded from the analysis in which evaluations are used as outcomes.

3.3.2 Variables used and descriptive statistics

School and student characteristics

Table 3.1 describes the public education system in Catalonia in the academic year 2013/2014, the last in our sample. There are 1556 primary schools and 538 middle schools. On average primary schools have 1.6 classes per grades, while middle schools have 3.3 classes per grade.[12]

About half of the students are female, 14.3% of students in primary education do not have Spanish citizenship (17.5% in lower secondary education).[13] We define three categories of parental education: low, average, and high. Parents' education is low if both parents have completed at most lower secondary education. It is high if one parent has tertiary education and the other has at least completed upper secondary education. It is average in the other cases.[14] About 36.4% of students in primary education has low educated parents, while 27.2% has highly educated parents. They are 41.8% and 22.0% respectively for lower secondary education.

5.5% of primary school students are "'behind", namely they are attending a lower grade than what normally expected for students of their age. The share is 8.7% among students who attend grade 6 (the last grade of primary school), while it raises to 16.2% in grade 7 (the first grade of primary school). Overall 22.6% of students in lower secondary education are behind. This suggests a sharp increase in grade repetition in middle school.

3.2% of students have special educational needs.[15] This assignment is quite persistent: both in primary and secondary education about 97% of students who are labeled as special needs in a given year have the same label the following period.

Evaluations

The main measure of performance used in the analysis is the average of the evaluations in the four core subjects that all students undertake throughout compulsory education: Mathematics, Catalan, Spanish, and English. In lower secondary education, evaluations are assigned using a 0-10 scale, while in primary evaluations the marks "Fail", "Pass", "Good", "Very good", "Excellent", are used. We convert them in numeric values using

[12]Statistics for students are similar in the previous years. However, as pointed out in section 3.3.1, not all the schools joined the data collection at the same time. In 2009/2010 only 717 primary schools appear in the sample (496 middle schools), their number sharply increases to 1440 in 2010/2011.

[13]For simplicity, from now on we will refer to students whose nationality is not the Spanish one as "immigrant". However some of them may be second generation immigrants, i.e. they may be born in Spain but not qualify for citizenship (e.g. if their parents just arrived in Spain).

[14]The classification is made using only the level of information of one parent when information for the other parent are missing.

[15]Special needs assignation and is regulated by Llei Orgànica d'Educació LOE 2/2006 (articles 71 and 72).

the same scale of evaluations assigned in secondary education.[16] Then we compute z-score at the grade–year–school level. Standardizing the evaluations within school serves the purpose of improving comparability across schools. In fact, tests are designed and graded by the teachers of the school, requirements to obtain a given score may vary substantially across schools. Figure 3.1 plots the distribution of GPA in primary education and secondary education.

Similarly, we compute the average score in Mathematics, Catalan, Spanish, and English for the region-wide test administered in grade 6 and 10. Then we compute z-score at the grade–year level. Internal and external evaluations are positively correlated. The correlation is 0.77 in grade 6 and 0.62 in grade 10.

School entry

Table 3.2 provides evidence that in all school years from 2009/2010 to 2013/2014 more than 99% of 6-year-old students are enrolled in the first grade of primary education. Moreover there isn't any evident trend that suggests an increasing attitude to postpone (or anticipate) entrance.[17]

We do not know the enrollment status at age 6 for children that were older when data collection started, therefore we cannot infer the share of non compliers for the previous school years. However, although we cannot assess the exact change due to the increased flexibility introduced by the decree issued in 2008, the overall effect of the change in law on enrollment behavior appears to be null or extremely small. Finally it is worth noting that some immigrant students may have started their education abroad in a country with different enrollment rules, both for the cutoff date and for the flexibility of the system. In fact, the share of delayed enrollment is 0.71% among natives and 1.7% among immigrants. We will study whether the age effect is different for native and immigrant students in Section 3.7.

Age at entry and outcomes

Figure 3.2 provides a first visual insight about the correlation between age at entry and educational outcomes in the short and longer run. The left panel plots the average GPA in second grade by month of birth: performances appear to be a linear decreasing function

[16]In middle school both numeric grades and the same five marks are assigned, thus each word correspond to an interval of numeric grades between 0 and 10. Using the same conversion scheme, we assign to each evaluation in primary school the midpoint of its interval; thus "Fail" is interpreted as 3, "Pass" as 5, "Good" as 6, "Very good" as 7.5 and "Excellent" as 9.5. An alternative approach would be to just use numbers from 1 to 5. If the analyses discussed next section are replicated using this second approach results are extremely similar.

[17]For postponed enrollment the share is slightly higher in 2012/2012, but the following year it drops back to 2011/2012 level. It is also worth noting that the statistics for 2009/2010 may be not fully comparable with the following years because the sample of schools involved is much smaller.

of the month of birth. On average students born in January perform more than 0.5 standard deviations better than their peers born in December. The right panel plots the empirical probability of undertaking academic upper secondary education by month of birth. On average younger children are less likely to enroll in academic high school than the older one (the difference between those born in January and those born in December is 5 percentage points), suggesting that the gap in maturity is persistent over time and has long run consequences.

A_i, the main measure of age at entry used in this paper, is based on students' day of birth. Days from 1 to 365 (or 366 in lap years) in the calendar year are mapped in the interval 0-1, so that January, 1 corresponds to 1 and December, 31 corresponds to 0. In fact the largest difference in age at entry for compliers is 1 year. As explained in previous subsection, almost all students comply with the rule of enrolling in primary school in the year in which they turn 6 years old. However a very small fraction, typically less than 1%, postpone the entrance. Therefore A_i, as inferred by student's i date of birth, can be regarded as the "expected entrance age" in primary school, rather than the actual entrance age. We can observe the actual entrance only for students born from 2003 to 2005, who start primary education from 2009 to 2013, but not for those born in previous years.

3.4 Empirical strategy

We study the effect of maturity at enrollment in primary school on children' achievement throughout compulsory education and on their educational choices at the end of lower secondary education.

When analyzing the age effect on a continuous outcome, we rely on the the following linear specification:

$$Y_i^t = \alpha^t A_i + X_i \gamma^t + \epsilon_i^t \tag{3.1}$$

Y_i^t is a measure of performance of student i in grade t, we use either the GPA assigned by teachers or the score in the region-wide test (for grade 6 and 10). X_i is a vector of time-invariant covariates realized before the child enroll in school, including dummies for gender, immigrant status, parental education, and cohort (i.e. calendar year of birth). α^t is consistently estimated under the assumption that $\mathrm{E}(\epsilon_i^t | A_i, X_i) = 0$. A_i, the age at enrollment in primary school, takes values in the interval 0-1, therefore the estimated coefficient $\widehat{\alpha^t}$ measures the difference in achievements between a student born on January 1 and a students born on December 31, everything else equal.

When analyzing the age effect on a binary outcome B_i, we estimate a probit model

$$\Pr(B_i = 1) = \Pr(\alpha_B A_i + X_i\delta + \varepsilon_i \geq 0) = \Phi(\alpha_B A_i + X_i\delta), \tag{3.2}$$

where Φ is the CDF of the standard normal distribution. The binary outcomes we study are grade repetition, dropout, enrollment in high school, and special needs diagnosis. Given that coefficients are only informative about the sign of the effect, we compute average marginal effects and marginal effects for relevant subsamples of the population.

The crucial assumption to correctly identify the age effect is that the age at entry A is uncorrelated with unobserved characteristics that affect the outcomes of interest. This assumption would fail if, for instance, parents who care more about the school performance of their children target the first weeks of the year to give birth. In subsection 3.4.2 we extensively discuss the plausibility of this assumption and we outline alternative specifications that we estimate as robustness checks.

If all students in our sample were observed when they start primary education, we could instrument the observed entrance age with the expected entrance age inferred from their date of birth. However, we observe most students for the first time in more advanced grades, and we only know their date of birth. Therefore we rely on the "reduced form" approach described by equations (3.1) and (3.2). Appendix 3.A compares results of the reduced form model with a two stage least square approach on the subset of students whose date of enrollment is directly observed. Given the extremely high rate of compliance the two approaches are virtually identical. For the sake of simplicity, in the rest of this paper we will refer to A_i as "age at enrollment" (or age at entry or entrance age), although we only observe the "expected" age.

3.4.1 Conceptual framework

As Elder and Lubotsky (2009) point out, entrance age may have lasting effects on human capital for two reasons. First, all children in a class are exposed to the same educational methods and contents, but they may have different learning capabilities due to different levels of maturity. Second, even if at some point their current production functions are identical, their level of human capital may be different due to their past history, and therefore they may end up with different human capital in the following period. To clarify this point, let A be the age at enrollment, Z^t other variables that contribute to human capital formation in grade t (for instance parental investment). Abstracting for now from grade repetition, let consider the following simple model of human capital accumulation:

$$\begin{aligned} H^t &= \beta H^{t-1} + \left(a^t A + Z^t b^t\right) \\ H^1 &= a^1 A + Z^1 b^1, \end{aligned} \tag{3.3}$$

H^t, the human capital in period t, depends on past human capital and current inputs Z^t. If $a^t > 0$, maturity has *direct* effect on human capital accumulation in grade t. In other words we can say that children born in different months have access to different technology to produce human capital. If $\beta > 0$, then there will be an *indirect effect* of producing differential levels of human capital through the accumulation process. In particular, if $a^t = 0$, A has only an *indirect* effect on H^t through past human capital H^{t-1}. Replacing H^{t-1} backward in (3.3) we obtain the following equivalent expression:

$$H^t = \left(a^t + \sum_{k=1}^{t-1} \beta^{t-k} a^k\right) A + \left(\sum_{k=1}^{t} \beta^{t-k} Z^k b^k\right), \tag{3.4}$$

A test score Y^t is a noisy measure of current level of human capital in grade t, i.e. $Y^t = H^t + \epsilon$. In practice in the empirical analysis we rely on individual and family characteristics which are constant over time as proxy for inputs to human capital, bringing to the data the empirical specification in equation (3.1). The estimated coefficients capture the cumulative effects of the observed characteristics over time, in particular $\alpha^t = \left(a^t + \sum_{k=0}^{t-1} \beta^{t-k} a^k\right)$ measures the *cumulative* effect of A on human capital in period t. We cannot disentangle the contemporaneous direct effect and the indirect effect accumulated over time, because β is unknown. However a comparison of the estimated coefficients across grades is informative of the prevailing effect: a decreasing sequence of $\widehat{\alpha^t}$ suggests that the main channel in the long run is the *indirect* effect, and that the direct effect is only relevant in the initial years. In particular, if β is sizeable, then a^t must be almost insignificant in the future, only the past accumulation process being relevant in explaining the long run gap.[18]

This simple illustration ignores grade repetition. Students who repeat a given grade may improve their stock of human capital at the end of that grade, and they are one year older when starting the following one. Grade repetition may counterbalance to some extent the effect of maturity, if younger students are more likely to be retained. We discuss this issue in subsection 3.5.2.

[18] Several works of Cunha, Heckman and other co-authors study life cycle skill formation, showing that skill attainment at one stage of the life cycle raises skill attainment at later stages of the life cycle. Their finding support the hypothesis that β is positive and sizeable. See Cunha et al. (2006) for a review.

3.4.2 Identification

As explained in section 3.5, the identification of the effect of age at enrollment A on student outcomes relies on the assumption that A is uncorrelated with unobserved characteristics that affect the outcomes. This assumption would not hold if, for instance, high SES parents prefer to give birth in the first months of the year, or low SES parents are more likely to deliver at the end of the year. In the analyses we control for parent's education, but this might not be enough if parents sort according to dimensions that are not perfectly captured by their education.

In the first part of this subsection we show that students are equally likely to be born at the beginning or at the end of the year, i.e. that there is no "manipulation" around the cutoff date of January, 1. Moreover we show that there is no difference in observable characteristics around the cutoff. In the second part of this subsection we discuss issues related to birth seasonality. In fact, if mothers with certain characteristics are more likely to give birth in specific periods the year, the estimated effects of age at entry may be confounded by birth seasonality even if there is no manipulation in a neighborhood of the cutoff date.

Distribution of births around the cutoff date

In the absence of manipulation, there should be no kinks in the density of the day of birth around the cutoff. We then formally test the no discontinuity hypothesis using three tests: the McCrary (2008) test, the Calonico, Cattaneo, and Titiunik (2014) test, and the Frandsen (2017) test. To perform those tests, cohorts are defined as running from July 1 of a given year to June 30 of the following year. The running variable is the day of birth, with January 1 taking value 0, December 31 taking value -1 and so on. We perform the tests on the sample of students born from July 2002 to June 2005, who are observed at age 6, 7, and 8 in our data.[19] Table 3.3 shows the p-values of the three tests on the entire sample and on the three subgroups defined by parental education. All the p-values are quite large and the null hypothesis of no discontinuity is never rejected. Figure 3.3 plots the data and the estimated density and confidence interval using the McCrary (2008) algorithm, providing further visual evidence that there is no discontinuity around the cutoff date.

Next, we provide evidence that the predetermined covariates (gender, immigrant status, parental education) are continuous around the cutoff date. First, we test whether the mean values are significantly different for students born around the cutoff date (more specifically in the last 15 days of December and in the first 15 days of January). As shown in column (1) of Table 3.4 we cannot reject the null hypothesis that there is no difference

[19]Performing the battery of tests on older cohorts of students produce similar results

for any of the variable under analysis. Second, we use students born throughout the year to test whether any of the covariates is discontinuous around the cutoff date. More specifically, for each predetermined covariate x_i we implement a regression discontinuity design using the following equation:

$$x_i = \zeta D_i + \theta f(d_i) + \kappa D_i f(x_i) + c_i + \epsilon_i, \tag{3.5}$$

where D_i is a dummy that take value 1 if the student is born between January 1 and June 30, $f(d_i)$ is a polynomial function of the day of birth, and c_i is the cohort fixed effect. We estimate (3.5) using the non parametric approach with a triangular kernel and a second order polynomial of the running variable, following the implementation proposed in Calonico, Cattaneo, and Titiunik (2014).[20] The parameter ζ captures difference in x_i before and after the cutoff. We report estimate of ζ and standard errors in column (2) of Table 3.4. None of the covariates exhibits a statistically significant difference around the cutoff.

A further concern that may arise is that some parents may take in account the day of birth of their child when choosing between public or private schools. For instance, they may think that a private school would suit better the needs of a child born in December. This type of parents would be over represented in January and under represented in December in our sample of children in public schools. We use enrollment data to test whether the probability of being enrolled in a public primary school is different before and after the cutoff date. Both the simple difference in mean and the RDD approach cannot reject the null hypothesis that there are no difference. We cannot reject null hypotheses also when the sample for the tests is restricted to students with a given parental education.[21]

Birth seasonality and parental background

Results discussed so far support the hypothesis that children born around the cutoff date are not different at the beginning of their life. However, the distribution of births throughout the year may differ by parental background. In column (3) of Table 3.4 we test whether the mean values of the covariates are different for children born from July to December and those born from January to June. We find that the latter are 1.6 percentage points more likely to have highly educated parents and 1.2 percentage points more likely to be native Spanish.

[20] Using a linear or higher order polynomial produces comparable results.

[21] We use 8-year-old students to perform the tests to make sure that we are taking in account all children, also the few children who delayed the entrance. The p-values of tests as the one in columns (1) and (2) of Table 3.4 are 0.3 and 0.6 respectively. For children of highly educated parents, they are 0.61 and 0.44 respectively.

Equations (3.1) and (3.2) are linear in age at entry A. If students born in the first half of the year are more likely to have good educational outcomes because of their parental background, the estimated coefficients of age at entry A in equations (3.1) and (3.2) may be upward biased. Figure 3.4 shows suggestive evidence that this should not be a concern. For the left panel of the figure, we regress a dummy for highly educated parents on a vector of dummies for the month of birth, controlling for gender, immigration status, and cohort. The graph plots the estimated coefficients with their confidence interval, showing that children born in March, April, and May, are more likely to have highly educated parents. The opposite is true for those born in August, while there are no significant difference for the other months. If differences in educational outcomes between older and younger students are due to unobservable correlated with their parental background, we would expect to see a similar pattern when we regress their GPA on the dummies for the month of birth. Conversely, as shown in the right panel of the figure, the GPA is decreasing in the month of birth in a very linear way.

The analyses in Section 3.5 rely on the assumption that A has a linear effect on evaluations (Equation 3.1) and on the latent variables of the binary outcomes studied with the probit model (Equation 3.2). In Section 3.6, we discuss two alternative approaches to support our finding. First, we replicate all the analysis using more flexible specifications. More specifically, we use dummies for the month of birth, or dummies for the week of birth.

Second, we replicate the analyses using only for students born at the beginning of January and at the end of December, who have similar observable characteristics as shown in Table 3.4. We compare the estimated coefficients for an indicator for being born in January with the estimated coefficients from our baseline specifications. If the estimated effects are spuriously capturing unobserved differences between children born in different periods of the year, we would find much smaller effects when using the subsample of students born around the cutoff date. It is important to remark that in this robustness check we are comparing outcomes of students born in January of a given calendar year with students born in December of the same year. We do not implement a regression discontinuity design, because it would require to compare students born in December of a given calendar year with students born in January of the following calendar year. However those students belong to different cohorts: they begin their education one year apart, they have different teachers, sit different exams, and are potentially exposed to different educational reforms. Those differences may bias in unpredictable ways any estimation.

3.5 Main results

3.5.1 Retention

We start our analysis by studying the effect of age at enrollment on probability of being retained during compulsory education. Results are shown in table 3.5. Each column shows results of a probit specification that follows equation (3.2). We exploit two types of dependent variables. First, we use dummy variables that take value 1 if the student has been retained in any of the first two grades of primary education (first column), and in any of the first two grades of lower secondary education (second column). Studying the probability of retention in the following grades would be less informative, because, as explained in Section 3.2, students can be retained at most once during primary education and at most twice during lower secondary education. We exploit only cohorts for which we can observe both students who are behind and students that are progressing at the regular pace up to the second grade of primary or lower secondary education. Therefore the sample for the analysis in the first column consists of students born from 2002 to 2005, and the sample for the second column consists of students born from 1997 to 1998.

Second, we assess the overall effect of maturity on the probability of being enrolled in a lower grade than the expected one given the year of birth. More specifically, we use dummies that take values 1 if the student is behind at 8 years old (i.e. she is not yet in grade 3), if the student is behind at 12 (i.e. she did not start secondary education yet), and if she is behind at 14 (i.e. she is not yet in grade 9). Results are shown in the last three columns of table 3.5. The chosen dummies are a close proxy for the event of having experienced retention at least once in the previous years, with the caveat that a student who started primary school late would be behind even if she was never retained during compulsory education.[22] In the analysis we exploit all observations belonging to students of that age in years 2009-2013. For instance we use students born from 2001 to 2005 to study the probability of not having reached grade 3 at 8 years old.[23]

On average, being one year younger increases probability of retention during the first two grades of primary school by 4.3 percentage points, a huge effect given that the average retention rate is about 3%.[24] Children who were younger when starting primary school are also more likely to be retained several years later during lower secondary education. The average marginal effect of A on the probability of being retained in the first two

[22] Ideally we would directly measure whether they have been retained at any point in time during the previous years, but we would need a longer panel to do so.

[23] Results do not change if we restrict the sample in the third column to data used in the first column, and the sample for the fifth column to the data used in the second column.

[24] The age effect is significant and sizeable, although smaller in size, also using as dependent variable retention during second and third cycle of primary school. Even if children born in the last months of the year are over represented among those retained in the first cycle, it is still more likely for younger children to be retained throughout primary school.

grades of middle school is -3.8 percentage points (16.5% of students are retained)

The remaining columns of table 3.5 confirm that younger students are significantly less likely to progress regularly throughout compulsory education. More specifically, the average marginal effect of A on the probability of being behind at 8 years old is -5.9 p.p. (4.3% of students are behind). Similarly, the average marginal effect on the probability of being behind at 12 (i.e. attending primary school rather than middle school) is 8.8 p.p. (9.2 % of students are behind), and the one on the probability of being behind at 14 is 10.9 p.p. (24.8% of students are behind). Most of these sizeable effects should be attributed to increased probability of retention during compulsory education for younger students. In fact, as discussed in section 3.3.2, overall less than 0.9% of students start primary education later, and the share is at most 2% in December.

3.5.2 Evaluations

We study the effect of age at enrollment on academic performance, from second grade (in primary education) to tenth grade (in secondary education). Subsection 3.5.2 describes the subsamples used and the implementation of the analysis, Subsection 3.5.2 discusses the results, Subsection 3.5.2 estimates quantile regressions to explore the effect of age at entry across the distribution of performance.

Sample selection

Performance are measured by internal evaluations in primary education (grades 2, 4, and 6) and lower secondary education (grades 7 to 10), and external evaluations at the end of each stage (grades 6 and 10). Given the limited time span of our sample we cannot follow the same pool of students throughout compulsory education, but we rather use different cohorts of students to estimate the maturity effect α^t for a given grade t.[25]

When analyzing internal and external evaluations in grade 10 we have a selection problem: children aged 16 can drop out of school without completing lower secondary education. Given the retention rules discussed in previous section, all children conclude at least the grade 8, but they may dropout before concluding grade 10 or even grade 9. This means that the sample of students for which we can observe evaluations in the two last years of middle school is not fully representative of the initial population of children. This selection issue is discussed extensively in section 3.5.3, where we analyze the effect of maturity on dropout. We include our results for grades 9 and 10 in the

[25] As a robustness check we restrict the estimation to students that we observe in two consecutive levels; for instance for those born in 2003 or 2004 we have evaluations both in grade 2 and in grade 4. We can then perform the relevant regressions only on this subsample and compare the results with those obtained with the general specification. This simple robustness check always confirms the validity of our approach. Results are available upon requests.

analysis discussed thereafter, in order to provide some illustrative evidence of the long lasting effect of maturity on performance, but we acknowledge that we cannot exclude either positive or negative biases due to selection.

The subsample used for each regressions includes only cohorts for which we can observe (at least) students with a regular progression or one year behind in primary education, and up to two years behind in secondary education. Therefore, we use four cohorts of students to study performance in primary education, and three cohorts for lower secondary education. For instance, when we study outcomes in second grade of primary school, we use students born from 2002 to 2005, while we do not include those who were born in 2001 (we would observe only students who are one or more year behind, but not those with a regular progression) or born in 2006 (we would observe only students in a regular progression, but not those retained the year before).[26] This choice is aligned with the regulation of enrollment and progression in primary and secondary school: students can be retained at most once during the entire cycle of primary school, and at most twice in secondary school. The region-wide test in grade 10 was introduced in the school year 2011/2012, therefore we use only the cohort of students bborn in 1996 when external evaluations in lower secondary education are the dependent variable.

Moreover, when a student undertakes the same grade twice only the most recent evaluation is included in the analysis. Consider for instance a student who attended the second grade of primary school in year 2010, was retained and repeated the same grade in 2011, being finally promoted. The evaluation obtained in 2011 is included, while the other is discarded. This approach ensures that we use the evaluation that the student obtained when she is admitted to the following grade.[27]

Being retained affect performance both in the grade repeated twice and in the following periods. Retained students have an additional year to learn the material covered in a given grade, and they are also one year older when they retake courses and exams. For instance, about 85% of students improve their GPA when they retake a grade in primary school. Moreover students who repeated grade t are older when they enroll in grade $t+1$ and, if retention is effective, they have a better stock of human capital than otherwise identical peers who did not repeat grade t. As shown in section 3.5.1 younger students are more likely to be retained, thus retention may contribute to close the gap due to the

[26] While for the oldest cohorts we can observe also students who are more than one year behind, this is not possible for 2005 cohorts. Given that the number of students who are two or more year behind or one year in advance is negligible, we ignore this difference. Further reducing the sample leaves the estimates unchanged

[27] When the final observation coincides with student's last year in the sample we do not have direct evidence that the student was promoted; however, given the choice of the sample, the regulation ensures that the student has to be promoted, except for very special cases. As a robustness check (not shown) we performed the same set of analysis exploiting one less cohort and imposing the further restriction that students are included in the sample only if we can observe them in a higher grade the following year. Results were completely equivalent to those discussed below.

day of birth. In this case, α^t in equation (3.1) is smaller than what would be found in an institutional setting without retention. Although such coefficients capture a combination of the effect of maturity on performance and on probability of being retained, results are still informative about the overall effect of the age at enrollment in primary school on the sequence of students' evaluations in a system that allows grade repetition.

Precisely because our primary interest is to understand the longer term consequences of the initial maturity gap, in our baseline specification we use the last available outcome for retained children. That is, if a child has been retained in sixth grade we analyze her evaluations the second time she takes the exams in sixth grade. We also replicate the analysis using evaluations the first time students take the exams. Comparison of the results give us a sense of how much retention help closing the gap between older and younger students in a given grade.

Results and alternative specifications

Each column of Table 3.6 (panel A) contains the estimated coefficients of a regression of a measure of performance on age A, individual characteristics, and cohort fixed effects, as in equation (3.1). For comparison, panel B reports results of a univariate regression of each dependent variable on A.

Table 3.6 provides compelling evidence that maturity is a very important determinant of students' performance at the beginning of their school career: *ceteribus paribus* being born at the beginning of January rather than at the end of December increases the GPA of 0.57 standard deviations. This effect is much larger than both the gender gap and the native-immigrant gap. It is just slightly smaller than the effect of having parents with tertiary education rather than with basic education.

The age effect is highly persistent over time, although decreasing in magnitude: it is 0.41 standard deviations in grade 4, and 0.33 s.d. in the last grade of primary education, while during lower secondary education being one year older is still associated with about 0.2 standard deviations increase in the GPA. Although the worst performing students, who (as we showed) are disproportionately born in the last months of the years, may have drop out before reaching the fourth grade, age effect is still significant at 1% level in grade 10.

Columns (ext.) confirm that results are not driven by the particular evaluation procedure used in schools. The estimated effect is similar and if anything slightly larger using the test scores obtained by the students in their external evaluations.

A comparison of the estimated coefficients in panel A with those in panel B suggests that the age effect is orthogonal to socio-economics. In fact the estimates are almost unchanged when additional regressors are introduced.

In a cross country analysis that exploits TIMSS and ECLS test scores, Bedard and Dhuey (2006) find that the effect of being born in January rather than in December range from 0.12 to 0.35 standard deviations in grade 4 and from 0.08 to 0.26 standard deviations in grade 8. Spain is not among the countries in the study. Our results for Catalonia suggest that the age at entry effect for Spain is among the highest in the world.

We replicate the analysis in table 3.6 separately by subjects, i.e. using as dependent variable the test score in Mathematics, Spanish, Catalan, or English. Results using the subjects rather than their GPA mimic the previous finding. Table 3.12 in the appendix shows results for internal evaluations in grades 2, 6, and 7.[28]

We augment the specification for each grade with peer characteristics, including average age. We focus on primary education, because some middle schools may sort children across classes based on their performance.[29] Results are shown in Table 3.13. Peers' average age, the share of highly educated parents, and the share of females affect negatively the internal evaluations, while they have a positive effects on external evaluations in grade 6.[30] To the extent that older students, students with high SES, and girls do better in school, these results are coherent with finding in Calsamiglia and Loviglio (2019), who show that being with better performing peers harm non-blind evaluations assigned by teachers. More importantly for the purpose of this paper, coefficients of own regressors are virtually unaffected by the introduction of the new regressors, confirming the robustness of the baseline specification.

Finally, we replicate the analysis using evaluations the first time students take the exams. Results, reported in Table 3.14, are quite similar to to those in Table 3.6. Not surprisingly the estimated $\widehat{\alpha}^t$ are slightly larger in magnitude, but the increase is always smaller than 4% of the baseline estimate. For instance the estimated age effect in grade 2 is 0.59 rather than 0.57, in grade 8 it is 0.23 rather than 0.22.

The fact that the effect of maturity on school outcomes decreases over time supports the hypothesis that younger children create a lower stock of human capital in the earlier stage of their academic career, but later on they do not cumulate human capital at a lower rate for a given level of human capital from previous period (otherwise the gap should be increasing rather than decreasing). However the initial disadvantage is so large that the negative effects propagate over time and the gap is not closed at the end of lower

[28] In the interest of space we do not report results for other grades. They are available upon requests.

[29] See Calsamiglia and Loviglio (2019) for a more detailed discussion of children's allocation to classes in primary and middle schools in Catalonia. We also replicate the analysis using peers at the school-year level, rather than at the class level, for both primary and secondary education, finding very similar results.

[30] Coefficients are always significant but the magnitude of the effects is relatively small compared with the effect of own characteristics. In fact, variation of peers' characteristics is relatively small within school. For instance an increase of 0.1 in peers' age at entry (which is much larger than one standard deviation) would decrease the GPA in second grade by 0.05 s.d., about one third of the effect of increasing individual age by 3 months (i.e. about one standard deviation in expected age).

secondary education.

A contrasting explanation is that the direct effect of maturity is still large in advanced grades, but the decreasing sequence of estimated coefficients is due to the higher number of younger kids who repeated a grade, partially closing the gap. While we cannot directly test the two alternative explanations, the latter appears unlikely for a number of reasons. First, we observe a clear decrease of $\widehat{\alpha^t}$ also during primary school, when the retention rate is relatively low. Second, as just discussed, the improvement of evaluations due to retention in a given grade matter little for the estimates. Third, a large share of students perform well enough to be basically unaffected by the possibility of retention. In next subsection we perform quantile analysis to investigate the age effect at different points of the distribution of evaluations. The same evolution of the age effect over time is found for all the quantiles.

Quantile analysis

We perform quantile regressions to study how the effect of age at enrollment change along the distribution of evaluations. We regress evaluations from grade 2 to grade 10 on age at enrollment A and the other covariates at various quantile. Figure 3.5 plots the estimated coefficients for the effect of age at enrollment on internal evaluations at the 25 percentile and 75 percentile. The evolution of the age effect over time is remarkably similar for the lower tail and the upper tail of the evaluations. As already observed for the linear regression model, the age effect is very large in grade 2, and it slowly decreases in the following grade, being still sizeable and significant at the end of lower secondary education. This pattern is similar at other quantiles, as shown in table 3.15.

For each grade and type of evaluations, the age effect is sizeable and significant throughout the entire distribution. This is an important confirmation that we should be concerned about the negative effect of being younger at the time of enrollment for all children, not only for those who exhibit particular characteristics.

The age effect is somewhat increasing in the quantiles of internal evaluations, with a small drop at the very top of the distribution in primary school. Conversely, it is decreasing when external evaluations are used as outcome.[31]

Difference in the results for internal and external evaluations may be due to the fact that the former are more suitable to assess student performance at the top of the distribution, while the latter are more suitable to capture difference in performance at the bottom. External evaluations are meant to test basic knowledge, while internal evaluations may allow to discriminate better the middle-high part of the distribution rather than the lower

[31] This result is in line with Arellano and Weidner (2016). Using as outcome the results in TIMSS (Trends in International Mathematics and Science Study) for a sample of Canadian students, they find that age effects are decreasing in ability.

tail. In fact only one grade is available for insufficient performance in primary school; in middle school in principle numbers from 1 to 4 can be used, but 3 and 4 are much more frequent in practice.[32]

3.5.3 Dropout and enrollment in academic high school

As discussed above, the age effect is persistent and still sizeable in middle school, but the trend is somehow decreasing over time. However, starting from the the year in which they turn 16, students have to take decisions that will have long lasting consequences on their labor market outcomes. First, if they turn 16 before graduation they can dropout without concluding lower secondary education. Second, after completing lower secondary education, they can enroll in further education, of either academic or vocational type. It is of uttermost importance to understand if the maturity effect is strong enough to affect their decisions: if the the younger children make on average different choices than their elder counterparts, it becomes evident that the initial gap in age has long term consequences. Thus our goal in this section is to explore whether age at the beginning or primary school affects the probability of completing lower secondary education, and then the probability of enrolling in the academic track of high school, and the probability of choosing vocational education.

We can expect two opposed effects on probability of graduation. To the extent that low performers are more likely to dropout, we can expect age to have a positive effect on the probability of graduation. On the other hand, older students face a larger time period in which they can dropout therefore they may have higher incentives to leave school before graduation. Which effect prevails is an empirical question.

The limited time span covered by our data imposes some restrictions on the sample used for the current analysis. We focus on students born in 1995 and enrolled in a public middle school in 2009, and determine whether they graduate from a public school in the following years, and whether they enroll in higher education.[33]

[32]However, it is noteworthy that the increasing and decreasing effects, respectively, are very robust when performing censored quantile regressions that consider bottom or top outcomes as censored.

[33]More precisely, we are identifying the effect of age at entry on the probability of concluding lower secondary education within the public system. We cannot follow students when they leave the public school and obtain a diploma through a private institution. However we do observe enrollment in both public and private high school. We do not find graduation for 371 students that are then registered in the academic track (1% of the sample under analysis in this section). Given that middle school diploma is a compulsory requirement for enrollment in the academic high school they should have obtained it in some unobservable way. We also observe 1722 students who enroll in vocational training without having completed lower secondary education in the public system. However this fact does not provide sufficient evidence that they obtained a middle school diploma, because, after turning 17, students who previously dropout have the possibility to access vocational education after the successful completion of some preparatory courses (another event that we cannot observe in the data). Given that counting as graduated only the students enrolled in batxillerat would slightly bias the results in our favor (older children are more likely to perform better and enroll in the academic track), we do not incorporate this

Although all children should attend school for at least some months before they turn 16 and can dropout, they may not be included in the records if schools transfer their data to the central system at the end rather than at the beginning of the years. Indeed while at earlier grades the probability of disappearing from the data in the following year is orthogonal with age, in the year in which they turn 14 children born at the beginning of the year are significantly more likely to disappear. Given this selection issue, to avoid to overestimate the age effect we analyze outcomes of 14 years old rather than 15 years old.[34]

73% of 14-year-old students enrolled in a public school attend grade 9, while 24.7% are one year behind and 2.3% two years behind. We classify them as "graduate" if they complete lower secondary education after experiencing retention for at most one additional time. In other words, a student enrolled in grade 9 in 2009 should finish in 2011 at the latest, while a student enrolled in grade 8 should finish in 2012 at the latest, and one enrolled in grade 7 should finish in 2013.[35] We adopt similar definitions for enrollment in academic or vocational upper secondary education.

Overall about 74% of students graduate, but students who are already behind at 14 years old are much more likely to dropout. In fact only 42.6% of those who are in grade 8 complete lower secondary education, and for those who lag two grades behind the graduation rate is as low as 16.1%. 50% of 14 years old eventually enroll in academic upper secondary education, while 24% enroll in vocational training. Only 13% of students who are one or more grade behind at 14 enroll in the academic track.

We regress each of the binary variables for the events of graduation, enrollment in further academic education, and enrollment in vocational training, on age A and covariates using the probit model described in equation (3.2). Estimated coefficients and marginal effects are reported in table 3.7.

Age at entry has a significant effect on the probability of gradution. The average marginal effect of A is 2 percentage points. This finding suggests that as far as dropout is concerned, the "negative" effect of being younger (due to average worst academic performance and increased probability of retention) completely offsets the "positive" effect due to longer time of compulsory education. This is in contrast with previous studies that exploit US data. The seminal work by Angrist and Keueger (1991) shows that younger children are more likely to stay in school. In a recent paper Cook and Kang (2016) find

information in our initial measure.

[34]Results we obtained replicating the analyses described in this section for 15 years old are qualitatively similar, but higher in magnitude. The coefficient of the regressor might spuriously capture the fact that the lower tail of distribution of ability of older children is not in the sample, thus older children in the sample had on average a better outcome. On the other hand our estimates may be downward biased because of higher measurement error: some of the 14 years old that we count as dropout may be truly leaving our sample of interest at 15 because moving out of the Catalan education system.

[35]We adopt this definition to allow the same delay for students in different grades at 14.

that, although older children obtain on average better evaluations before turning 16, they are then more likely to dropout and be engaged in criminal activities when adult. The peculiarity of the Spanish system, in which students who stay in school can achieve an official qualification a few months after they turn 16, may explain part of the difference in results: older children have more incentives to stay in school, therefore the "negative" effect of being younger is prevalent in our analysis.
Remaining columns of table 3.7 show that maturity has a sizeable and significant effect on the probability of enrolling in further academic education: on average being one year older increases the probability of enrolling in further academic education by 5.6 percentage points, and decreases the probability of enrolling in vocational training by 3.3 percentage points.

3.5.4 Diagnosis of learning disabilities

In public schools children with special education needs may be granted additional support during compulsory education. As discussed in section 3.3.2, each academic year less than 4% of students in our data are labeled as "special needs" children. While physical disabilities should be straightforward to identify, behavioral or learning diseases (such as attention deficit or mild intellectual disability) are typically diagnosed while the student attends school, with teachers being instrumental in its detection and diagnose. After the condition is confirmed, the child is formally classified as needing special education arrangements.

Our concern here is that teachers' perceptions may be partially clouded by children's maturity: additional immaturity of a younger child may be confounded with a learning disability. Conversely, older children who would benefit from special support may be under diagnosed. We provide evidence in support of this hypothesis, testing whether age at entry A affects the probability of being labeled with a "special needs" code related to learning disorder (development and behavioral disorders, mild intellectual disability, and other conditions whose detection may require some subjective judgment of the educator). As placebo test we perform the same analysis using as dependent variable a dummy for physical disability (blindness, deafness, mobility issues, whose diagnosis should be relatively objective). Results are reported in table 3.8. The first two columns refer to the first two grades of primary school, while the other refer to first two grades of middle school. We use students born in years 2003 to 2005 for primary education and students born from 1997 to 1999 for lower secondary education.

Being one year older reduces significantly the probability of being diagnosed with learning disorder in the first grades of primary school, and the effect is quite stable over time. In fact the average marginal effect of age at entry is 1.6 percentage points in primary school and 1.5 percentage points in middle school. On the other hand age at entry has no effect on the probability of being attributed physical disability.[36]

3.6 Robustness checks

In the following subsections we discuss the results of the robustness checks presented in previous Section 3.4.2. Moreover, in Appendix 3.A we show that reduced form and instrument variable approach deliver the same results. In Appendix 3.B, we replicate some of the analyses discussed in Section 3.5 using students enrolled in private schools. Results are fully aligned with those for public schools, confirming that our findings hold for all students enrolled in the Catalan educational system.

3.6.1 Specifications flexible in the time of birth

We replicate all the analyses in Section 3.5 using dummies rather than one continuous variable for the age at entry. We use either dummies for the month of birth or dummies for the week of birth. For instance, we regress Y_i^t, the evaluations in grade t, on a vector of dummies for the month of birth and the usual covariates:

$$Y_i^t = \sum_{m=1}^{11} \alpha_m^t M_{m,i} + X_i \gamma^t + \epsilon_i^t, \tag{3.6}$$

where $M_{m,i}$ takes value 1 if the student i is born in month m. December is used as reference category, therefore α_m^t is the expected difference in evaluations between a student born in month m and an otherwise identical student born in December. We also run the same regression using dummies for the week of birth, week 52 is used as reference category.

Figure 3.6 plots the estimated coefficients and confidence intervals of the month dummies for GPA in grades 2, 6 and 7, retention in the first two grades of primary school, being behind at 14 years old, enrollment in high school. Each graphs displays also a linear fit of the estimated coefficients. The estimated coefficients for the January and November, and for the other covariates are reported in columns (M) of Table 3.16. Figure 3.7 replicates Figure 3.6 using weekly dummies, columns (W) of Table 3.16 show the coefficients.

Results using the month dummies confirm the findings discussed in previous Section 3.5. For each specification, the estimated effect of being born in January rather than in December is quite similar, only slightly smaller in magnitude, to the estimated effect of being one year older. Moreover all outcomes are decreasing in the month of birth. The estimated effects on evaluations throughout compulsory education almost perfectly lie on

[36]The share of the students with special needs related to learning is 2.8% in primary school and 3.2% in middle school. The corresponding figures for physical disabilities are 0.46% and 0.35% respectively.

their linear fit.[37] The pattern is remarkably similar for the more demanding specifications that use the week dummies.[38]

3.6.2 Analyses restricted to the subsample of students born around the cutoff

We replicate the analysis in Section 3.5 using only students born in the initial 15 days of January and in the last 15 days of December (16 for lap years). For instance, we regress Y_i^t, the evaluations in grade t, on a dummy old_i and the usual covariates:

$$Y_i^t = \alpha_{\text{old}}^t \text{old}_i + X_i \gamma^t + \epsilon_i^t, \tag{3.7}$$

where old_i takes value 1 if the student is born in January. α_{old}^t is the expected difference in evaluations in grade T between students born at the beginning and at the end of the year.

Results of the analyses on the much smaller sample of students born around the cutoff date are quite aligned with the results discussed in Section 3.5. More specifically, the estimated effect on evaluations are quite close to those found for the entire sample. For instance $\alpha^2 = 0.574$, while $\alpha_{\text{old}}^2 = 0.531$. The average marginal effects on the probability of being retained or being behind are the same or slightly larger in magnitude. The marginal effects on the probability of receiving a special needs diagnosis are also very similar to those found for the entire sample (e.g. 1.5 percentage points in primary school).

Estimated marginal effects on the probabilities of graduating and pursuing further studies are close enough. The average marginal effect on graduation is 3.5 percentage points (2 p.p. using the full sample), while the average marginal effect on enrollment in academic upper secondary education is 4.1 percentage points (5.6 p.p. for the full sample). The estimated marginal effect on enrollment in vocational education is -1 percentage point but it is not significant (it is -3.3 p.p. in the full sample). As explained in Section 3.5.3, those analyses are performed using only one cohort of students, those born in 1995, therefore the sample size is relatively small. This may explain why the estimates are less precise than the ones discussed above.

Table 3.17 shows the estimated coefficients and marginal effects for a subset of outcomes: GPA (internal evaluations) in grades 2, 6 and 7, retention in the first two grades

[37] The only exception are May and August in grade 7, which are somewhat higher and somewhat lower respectively. Given that highly educated parents are more likely to have children in May and less likely to have children in August the dummies for this two months may be slightly biased by unobservable related to parental background.

[38] In the interest of space we do not show results for the other regressions. They are aligned with the results described in Section 3.5. Results are available upon requests.

of primary school, being behind at 14 years old, enrollment in high school.[39]

3.7 Heterogeneity analysis

We study whether the effect of age at entry is heterogeneous across subgroups of the population. We augment the specifications estimated in Section 3.5 with interaction terms between age at entry and parental education, gender, immigrant status.

3.7.1 Retention and evaluations

As shown in Table 3.18, the effect of age on grade repetition or on the event of being behind does not exhibit significant difference by parents' type. However, as reported in Table 3.9, the marginal effects of age at entry on retention probability are decreasing in parental education. For instance, being one year younger increases the probability of being behind at age 14 of about 13 percentage points for students with low educated parents, while the increases for students with highly educated parents is only 6 percentage points. This is not surprising, given the large effect that parental education has on the outcome: in the probit model, any change affects more those whose latent variable is closer to 0, i.e. whose probability is closer to 50%. For instance, 34% of students with low educated parents are behind at 14 years old, the share are 18% and 8% for students with average and highly educated parents, respectively. Columns (1) and (2) of Table 3.9 confirm that the estimated marginal effects are almost identical for the baseline model and the model with interactions.[40] In primary education, the estimated age effect on internal evaluations is slightly larger for students whose parents have an average level of education. As shown in Table 3.18, the difference is significant in grade 2 and 4, but it is small in size (e.g. it is 0.05 s.d. in grade 2, when the effect of being one year older is 0.54 for students with low educated parents). There are no differences between students with low and high educated parents. Conversely, in lower secondary education, the age effect is about 0.07 s.d. larger for students with high educated parents, while it is not significantly different for those with average education level. On the other hand, results using external evaluations in grade 6 show that age affects slightly less the performance

[39]Results for the remaining specifications are available upon request.

[40]Our marginal effects are close enough to the estimates in Berniell and Estrada (2017). They use Spanish data from PISA tests and show that among low SES, students who are born in December are 12.7 p.p. more likely to experience retention during compulsory education, while among high SES the gap decreases to 4 p.p. However we think that this difference alone is not evidence of high SES investing more in their children if they are younger. In fact, as discussed, it can be observed also if high SES invest the same in children born in December and in children born in January, but they are on average more likely to have high performing offspring.

of students with highly educated parents.[41]

The age effect is also similar for male and female. The effect on retention is slightly smaller in size for female during primary education, while there are no significant difference in secondary education. The marginal effects are smaller for girls, who on average are less likely to experience retention. When evaluations are used as outcome, the estimated coefficients of the interaction term are positive but small in size, and only significant in lower secondary education.

Overall, the differences in the age effect on retention and evaluations by parental education or gender are small, if any. The age at entry has large impact on educational outcomes throughout compulsory education for both students with high and low socioeconomic status (SES), and for both boys and girls.

This finding contrasts with Elder and Lubotsky (2009), who find that in US the effect is increasing in children's SES.[42] One possible explanation comes from the differences in preschool systems in Spain and US. Preschool quality is quite heterogeneous in US and a substantial fraction of children do not attend preschool regularly. Parental contribution is therefore fundamental to determine the human capital level of the child at the beginning of formal education. Younger children start school having spent less time with their parents, and *ceteris paribus* the gap is larger among high SES, given that they have foregone more parental inputs when starting school. Conversely in Spain almost every child is enrolled in preschool, for which there is free and universal access starting in the year in which the child turns 3. Even if the gap that we observe in primary school may be even larger in preschool, the quality of the time spent out of school before age 6 depends less on their SES.[43]

As shown in Table 3.18, the age effect on evaluations is significantly lower for immigrant students. The gap is larger in more advanced grades.[44] A plausible explanation for this result is that many non Spanish students are recent immigrants. They may have experienced a different educational system for the first part of their education, with different cutoff dates or simply more flexibility, and therefore they may be less affected by our measure of maturity at enrollment. Although this is a speculation we cannot directly test, it is supported by the evidence that the difference widen in more advanced grades,

[41] There are no significant differences across the three categories of parental education when external evaluations in grade 10 are used. However, as explained in Section 3.5.2, students self select into attending the last grade and therefore in taking the test. This can be a particularly relevant issue for the heterogeneity analysis because students with different parental background have different propensity to dropout. We do not report results for grades 9 and 10 in the table.

[42] According with their results, the age effect on test scores is more than double for high SES than for low SES throughout primary education.

[43] Younger children from high SES may foregone more inputs before the enrollment in preschool. However those inputs may matter less for their performance in school.

[44] Wald tests reject the null hypotheses that the effect is 0 for immigrants from grade 2 to grade 7. The null hypothesis cannot be rejected in grade 8.

where recent immigrants have been exposed to relatively more education abroad than in Spain.

3.7.2 Other outcomes

Table 3.19 replicates the analysis in Sections 3.5.3 and 3.5.4, including the interaction terms in the specification. Results point to a larger effect of age at entry on dropout and enrollment in high school for students in the middle category of parental education. Conversely there are no significant differences between students with low and high educated parents. For graduation, the estimated effect on the latent variable is almost double for students whose parents have average level of education. For this category, the estimated marginal effect on the probability of graduation is 4.4 percentage points, while it is 2 p.p. using the baseline model (Table 3.9). The coefficient of the interaction between age and the dummy for the middle category is also large and significant for the enrollment in further academic education. The estimated marginal effect is 9.6 percentage point, while it is 6 percentage point using the baseline specification. Similarly, there is a significant and sizable negative effect on enrollment in vocational training.

Completing lower secondary education and enrolling in high school are by a large extent a choice of the student and his or her family.[45] Our findings are compatible with the hypothesis that family with average SES are more responsive to the level of skills of their children when taking educational decisions. For instance, this would be the case if high SES always push their children to acquire more academic education, even if they are not performing well in school, while low SES do not push them enough, even if they have the potential to succeed.

As discussed in previous section, age at entry affect similarly performance of girls and boys. However, results in Table 3.19 suggest that age matters much more for boys than for girls for graduation and enrollment in academic upper secondary education. In fact, we cannot reject the null hypotheses that age at entry does not affect female graduation and enrollment.[46]. One possible reason for this results is that females are less responsive to their level of skills when choosing whether to dropout or to enroll in further education.

Results in the last two columns of Table 3.19 confirms that being younger increase the probability of being diagnosed with learning disabilities for all types of children. At the beginning of primary school the age effect is smaller for students with highly educated parents and it is larger for students whose parents have average educational level. However they are the same in more advanced grades.

[45] Conditional on completion of lower secondary education, enrollment in high school is free and guaranteed to all applicants.

[46] We test whether the sum of the coefficients of "Age" and "Age X female" are 0. The p-values are 0.71 for graduation and 0.14 for enrollment in high school

3.8 Conclusions

This paper shows that an inflexible system that allows for no postponing of the entrance of children and little retention early on may lead to maturity effects to have long lasting effects. We use a very rich data set including internal and external evaluations to confirm the persistence of the disadvantage for younger children over time and across the ability distribution. The effect is not only leading to worse performance over time, but to their choices as to whether to continue their education and what career to pursue to be significantly different. The fact that the effect is decreasing over time suggests that having a flexible system that better adapts individual maturity levels to specific grades early on would better adapt the initial heterogeneity, allowing for a better progression over time.

3.9 Tables

Table 3.1: Public schools in Catalonia

	Parents' education			Female	Immigrant	Behind	Special needs	Schools	
	Low	Avg	High					N	Classes
Primary	36.4%	36.5%	27.2%	48.4%	14.3%	5.5%	3.2%	1556	1.6
Middle	41.8%	36.2%	22.0%	48.6%	17.5%	22.6%	3.2%	538	3.3

Notes. Statistics computed for students enrolled in a public primary or middle school in Catalonia in the academic year 2013/2014.

Table 3.2: Delayed or early enrollment in primary education

School year	Share who ...		N students
2009/2010	*delays*	0.78%	
	anticipates	0.064%	23449
2010/2011	*delays*	0.84%	
	anticipates	0.045%	48980
2011/2012	*delays*	0.83%	
	anticipates	0.036%	52646
2012/2013	*delays*	0.92%	
	anticipates	0.056%	53432
2013/2014	*delays*	0.82%	
	anticipates	0.049%	53156
Total	*delays*	0.84%	
	anticipates	0.048%	231663

Notes. For each year the table shows the share of 6 years old who are still in preschool or are already in second grade. Last column shows the number of 6 years old in each year.

Table 3.3: Tests for continuity of the density of the day of birth

	All	Parents' edu high	Parents' edu avg	Parents' edu low
McCrary (2008)	0.889	0.656	0.348	0.947
Calonico et al (2014)	0.349	0.564	0.383	0.788
Frandsen (2017)	0.861	0.344	0.700	0.186

Note. Tests are performed using the sample of students born from July 2002 to June 2005. Each entry in the table is a p-value. We allow the McCrary (2008) test and the Calonico et al. (2014) test to select the optimal bandwidth independently. For the McCrary (2008) test we set a bin size of 1 day, to account for the discrete nature of our running variable.

Table 3.4: Tests for balance of predetermined covariates

	(1)	(2)	(3)
Female	0.000	-0.000	-0.001
	(0.010)	(0.014)	(0.003)
Immigrant	0.003	-0.001	0.012
	(0.007	(0.011)	(0.002)**
Parents' edu low	-0.006	0.003	0.019
	(0.010)	(0.015)	(0.003)**
Parents' edu avg	0.015	-0.014	0.001
	(0.010)	(0.015)	(0.003)
Parents' edu high	-0.002	-0.003	-0.016
	(0.009)	(0.012)	(0.003)**

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. Tests are performed using the sample of students born from July 2002 to June 2005. For every line of the table, each entry reports the estimated difference of the covariate for students born before and after the cutoff date (January 1). The standard error of the estimate is reported in parenthesis. Columns (1) and (3) perform a t-test for difference in mean; the sample for (1) includes only students born in the first 15 days of January and in the last 15 days of December, while all students are included in (3). Column (2) uses all the students in the sample, and perform a local non-parametric RDD specification following Calonico et al. (2014) (we allow the algorithm to select the optimal bandwidth).

Table 3.5: Grade repetition during compulsory education

	Repeat		Behind at		
	grade 1 or 2	grade 7 or 8	age 8	age 12	age 14
Age at entry	-0.637	-0.163	-0.698	-0.597	-0.382
	(0.023)**	(0.020)**	(0.019)**	(0.017)**	(0.012)**
	[-0.043]	[-0.038]	[-0.059]	[-0.088]	[-0.109]
	<0.002>	<0.005>	<0.002>	<0.004>	<0.003>
Female	-0.123	-0.308	-0.139	-0.143	-0.283
	(0.013)**	(0.013)**	(0.011)**	(0.009)**	(0.008)**
	[-0.008]	[-0.072]	[-0.012]	[-0.021]	[-0.081]
	<0.001>	<0.003>	<0.001>	<0.002>	<0.002>
Immigrant	0.327	0.226	0.444	0.672	0.523
	(0.020)**	(0.019)**	(0.017)**	(0.027)**	(0.015)**
	[0.022]	[0.053]	[0.037]	[0.099]	[0.149]
	<0.001>	<0.004>	<0.002>	<0.006>	<0.004>
Parents' edu avg	-0.365	-0.350	-0.328	-0.336	-0.397
	(0.016)**	(0.015)**	(0.014)**	(0.018)**	(0.011)**
	[-0.024]	[-0.081]	[-0.028]	[-0.050]	[-0.113]
	<0.001>	<0.004>	<0.001>	<0.003>	<0.003>
Parents' edu high	-0.704	-0.854	-0.596	-0.661	-0.899
	(0.023)**	(0.023)**	(0.020)**	(0.029)**	(0.018)**
	[-0.047]	[-0.199]	[-0.050]	[-0.098]	[-0.256]
	<0.002>	<0.006>	<0.002>	<0.006>	<0.005>
N	163,910	72,079	205,242	188,075	178,871

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. Cohort FE included. Standard errors in first and second columns ("Repeat...") are clustered by school. Standard errors in the other three columns ("Behind at...") are robust. Average marginal effects and their standard errors are reported in squared and angled brackets respectively.

Table 3.6: Evaluations in primary and lower secondary education

Panel A

	grade 2	grade 4	grade 6 (int.)	grade 6 (ext.)	grade 7	grade 8	grade 9	grade 10 (int.)	grade 10 (ext.)
Age at entry	0.574	0.407	0.329	0.351	0.220	0.175	0.153	0.137	0.219
	(0.008)**	(0.008)**	(0.009)**	(0.009)**	(0.010)**	(0.010)**	(0.010)**	(0.011)**	(0.021)**
Female	0.170	0.218	0.293	0.134	0.339	0.338	0.339	0.324	0.071
	(0.005)**	(0.005)**	(0.005)**	(0.006)**	(0.007)**	(0.007)**	(0.008)**	(0.008)**	(0.013)**
Immigrant	-0.204	-0.222	-0.227	-0.414	-0.243	-0.225	-0.248	-0.305	-0.742
	(0.011)**	(0.011)**	(0.011)**	(0.014)**	(0.014)**	(0.014)**	(0.013)**	(0.013)**	(0.025)**
Parents' edu avg	0.326	0.348	0.342	0.435	0.324	0.305	0.258	0.188	0.401
	(0.006)**	(0.007)**	(0.007)**	(0.010)**	(0.008)**	(0.009)**	(0.009)**	(0.008)**	(0.016)**
Parents' edu high	0.621	0.670	0.679	0.837	0.681	0.661	0.591	0.492	0.854
	(0.008)**	(0.008)**	(0.009)**	(0.012)**	(0.012)**	(0.012)**	(0.011)**	(0.012)**	(0.020)**
R^2	0.12	0.12	0.12	0.16	0.13	0.12	0.11	0.09	0.21
N	160,438	148,472	136,281	117,752	106,707	102,960	96,170	86,025	24,465

Panel B

	grade 2	grade 4	grade 6 (int.)	grade 6 (ext.)	grade 7	grade 8	grade 9	grade 10 (int.)	grade 10 (ext.)
Age at entry	0.583	0.412	0.334	0.350	0.220	0.170	0.154	0.144	0.226
	(0.009)**	(0.009)**	(0.009)**	(0.010)**	(0.010)**	(0.010)**	(0.011)**	(0.012)**	(0.023)**

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. In panel A, cohort FE are included and standard errors clustered by school. In panel B standard errors are robust.

Table 3.7: Probability of completing lower secondary education and undertaking further education

	Lower sec. edu.	Upper sec. edu.	
	Graduate	Academic	Vocational
Age at entry	0.068	0.162	-0.108
	(0.028)*	(0.025)**	(0.025)**
	[0.020]	[0.056]	[-0.033]
	<0.008>	<0.009>	<0.008>
Female	0.350	0.416	-0.246
	(0.017)**	(0.016)**	(0.016)**
	[0.104]	[0.144]	[-0.075]
	<0.005>	<0.005>	<0.005>
Immigrant	-0.558	-0.554	-0.145
	(0.024)**	(0.024)**	(0.023)**
	[-0.166]	[-0.192]	[-0.044]
	<0.007>	<0.008>	<0.007>
Parents' edu avg	0.363	0.503	-0.155
	(0.019)**	(0.017)**	(0.019)**
	[0.108]	[0.174]	[-0.047]
	<0.006>	<0.006>	<0.006>
Parents' edu high	0.747	1.110	-0.547
	(0.029)**	(0.025)**	(0.024)**
	[0.223]	[0.385]	[-0.166]
	<0.008>	<0.008>	<0.007>
N	33,624	33,624	33,624

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. Cohort FE included. Standard errors clustered by school. Average marginal effects and their standard errors are reported in squared and angled brackets respectively.

Table 3.8: Diagnosis of special needs

	Grades 1 and 2		Grades 7 and 8	
	learning	physical	learning	physical
Age at entry	-0.256	-0.024	-0.217	0.039
	(0.025)**	(0.045)	(0.025)**	(0.057)
	[-0.016]	[-0.000]	[-0.015]	[0.000]
	<0.002>	<0.001>	<0.002>	<0.001>
Female	-0.334	-0.053	-0.289	-0.011
	(0.015)**	(0.026)*	(0.016)**	(0.034)
	[-0.021]	[-0.001]	[-0.020]	[-0.000]
	<0.001>	<0.000>	<0.001>	<0.000>
Immigrant	0.006	0.024	0.069	0.036
	(0.028)	(0.039)	(0.022)**	(0.044)
	[0.000]	[0.000]	[0.005]	[0.000]
	<0.002>	<0.001>	<0.002>	<0.000>
Parents' edu avg	-0.213	-0.028	-0.361	-0.095
	(0.021)**	(0.032)	(0.018)**	(0.042)*
	[-0.013]	[-0.000]	[-0.025]	[-0.001]
	<0.001>	<0.000>	<0.001>	<0.000>
Parents' edu high	-0.391	-0.039	-0.589	-0.078
	(0.025)**	(0.035)	(0.027)**	(0.049)
	[-0.024]	[-0.001]	[-0.041]	[-0.001]
	<0.002>	<0.000>	<0.002>	<0.000>
N	149,142	149,142	115,204	115,204

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. Cohort FE included. Standard errors clustered by school. Average marginal effects and their standard errors are reported in squared and angled brackets respectively.

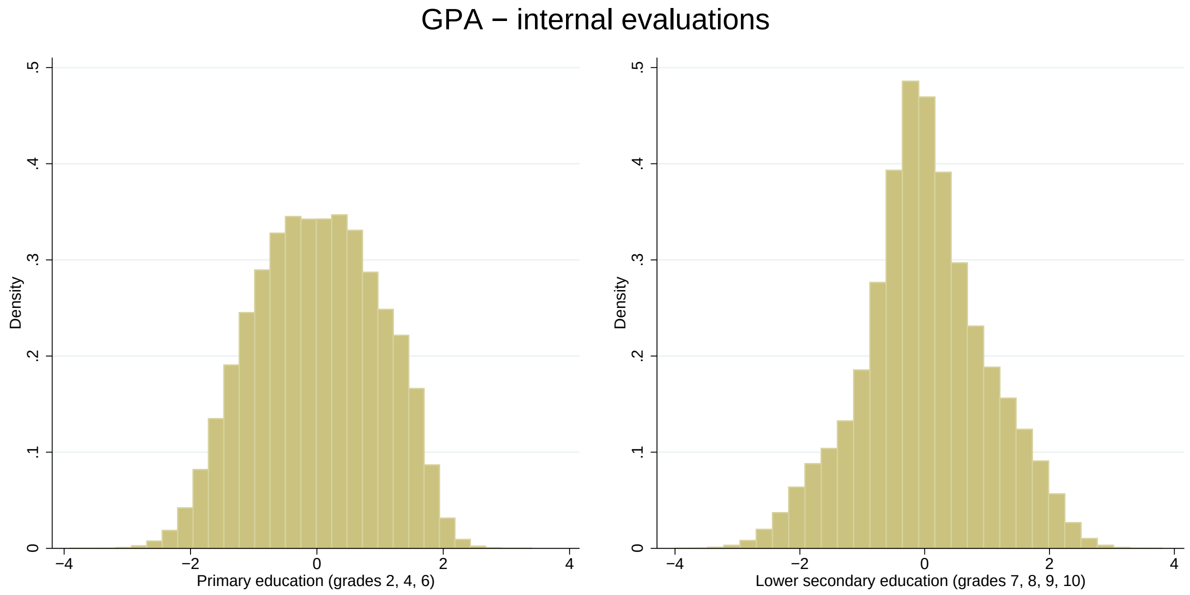
Table 3.9: Marginal effects of age at entry by parents' education

	Repeat grade 1 or 2		Behind at 14		Graduate		Enroll Academic	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Parents' edu. low	-0.066	-0.065	-0.134	-0.131	0.024	0.014	0.058	0.041
	(0.003)**	(0.004)**	(0.004)**	(0.006)**	(0.010)*	(0.012)	(0.009)**	(0.013)**
Parents' edu. avg	-0.035	-0.038	-0.106	-0.113	0.020	0.044	0.061	0.096
	(0.002)**	(0.003)**	(0.003)**	(0.006)**	(0.008)*	(0.014)**	(0.009)**	(0.015)**
Parents' edu. high	-0.017	-0.017	-0.064	-0.060	0.014	-0.003	0.048	0.024
	(0.001)**	(0.002)**	(0.002)**	(0.006)**	(0.006)*	(0.015)	(0.007)**	(0.019)
All	-0.043	-0.043	-0.109	-0.109	0.020	0.021	0.056	0.057
	(0.002)**	(0.002)**	(0.003)**	(0.003)**	(0.008)*	(0.008)*	(0.009)**	(0.008)**

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. The table plots the average marginal effects of age at entry when parents' education is set to be low, average, or high, and the overall average marginal effects (row "All"). Columns (1) use the baseline specifications estimated in Table 3.5 (for dependent variables "Repeat grade 1 or 2" and "Behind at 14") and Table 3.7 (for dependent variables "Graduate" and "Enroll Academic"). Columns (2) use the specifications augmented with interaction terms in Table 3.18 (for dependent variables "Graduate" and "Enroll Academic") and Table 3.19 (for dependent variables "Graduate" and "Enroll Academic").

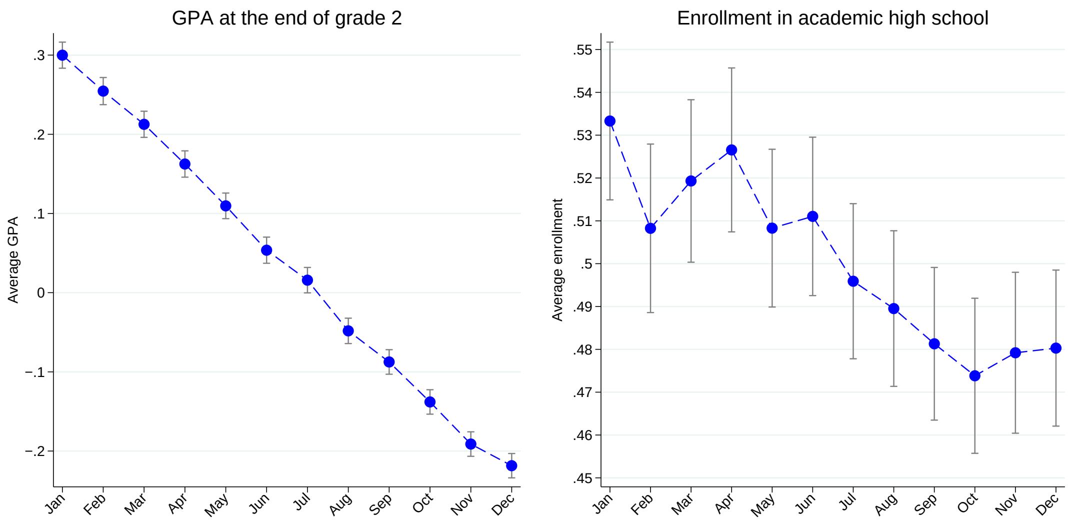
3.10 Figures

Figure 3.1: Distribution of evaluations



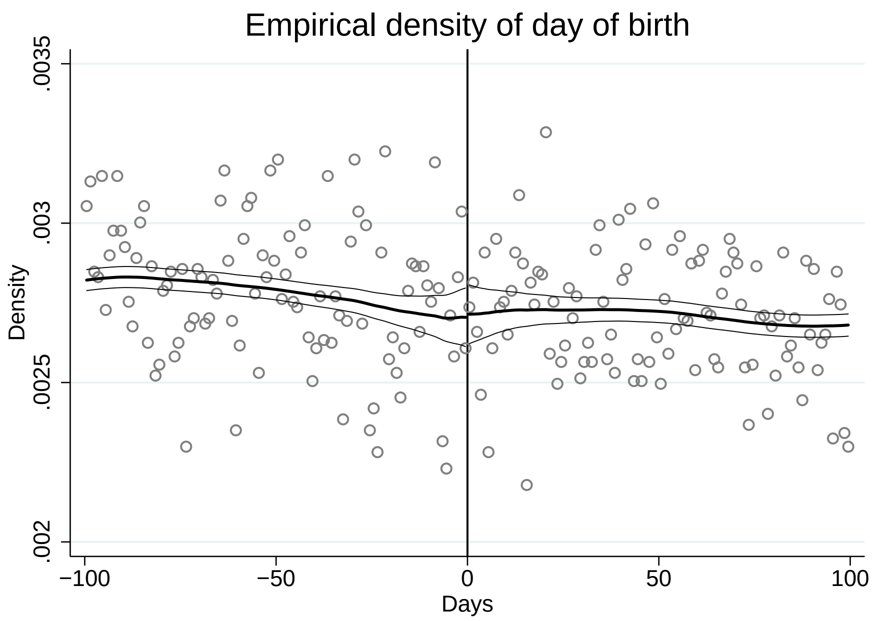
Distribution of GPA in primary schools (left panel) and middle schools (right panel).

Figure 3.2: Mean outcomes by month of birth



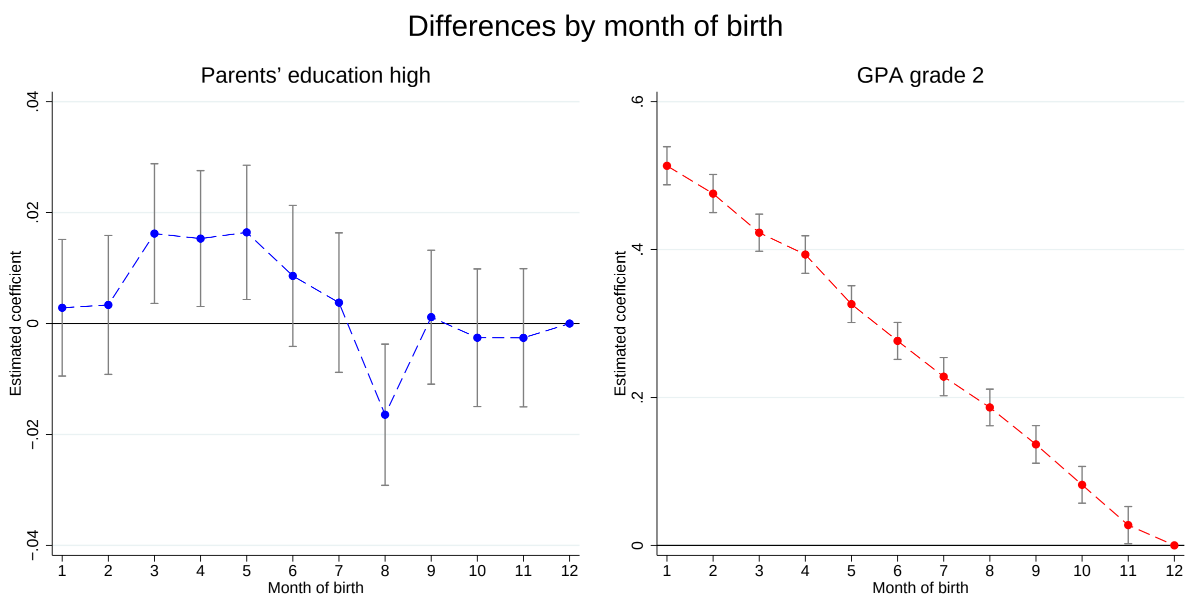
The left panel plots the average GPA at the end of second grade for students born from 2002 to 2005. The right panel plots the share of enrollment in academic high school among students born in 1995.

Figure 3.3: Distribution of births in the calendar year



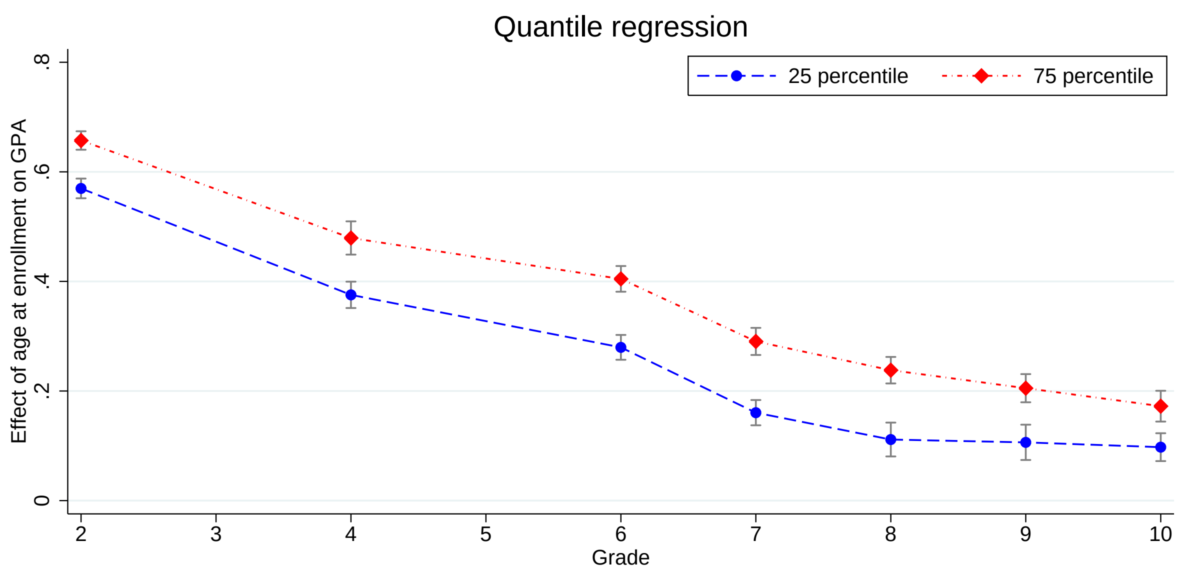
Distribution of births around the cutoff date of January 1 (day 0). The figure plots the empirical density and confidence band estimated using the McCrary (2008) algorithm, using students born from July 2002 to June 2005.

Figure 3.4: Correlation of month of birth with parents' education and student's performance



We regress a dummy for highly educated parents (left figure) or the GPA in grade 2 (right figure) on a vector of dummies for month of birth, controlling for gender, immigrant status and cohort. The figures plot the estimated coefficients with their confidence interval.

Figure 3.5: Effect of age at enrollment on GPA over time



The graph plots the results of quantile regressions of GPA (internal evaluations) on age at enrollment at the 0.25 and 0.75 quantile over school grades. Each blue dot is the estimated marginal effect for age at enrollment A at the 0.25 quantile in a given grade. Each red diamond is the estimated marginal effect for A at the 0.75 quantile. Grey bars are 95% confidence intervals

Appendix 3.A Two stage least square regressions

As explained in Section 3.5, we do not observe the first enrollment in primary education for most of the students in our sample. Therefore, we infer the enrollment year from the date of birth. However, a subset of students (those born between 2003 and 2005) is observed both when they begin primary school and when they are evaluated at the end of their second year. We exploit this subset of children to compare two stage least square estimates with our reduced form approach.

Using the same notation of Section 3.5, and calling OA_i the observed age at enrollment for individual i, the appropriate specification for evaluations Y_i as linear function of OA_i and other covariates X_i is:

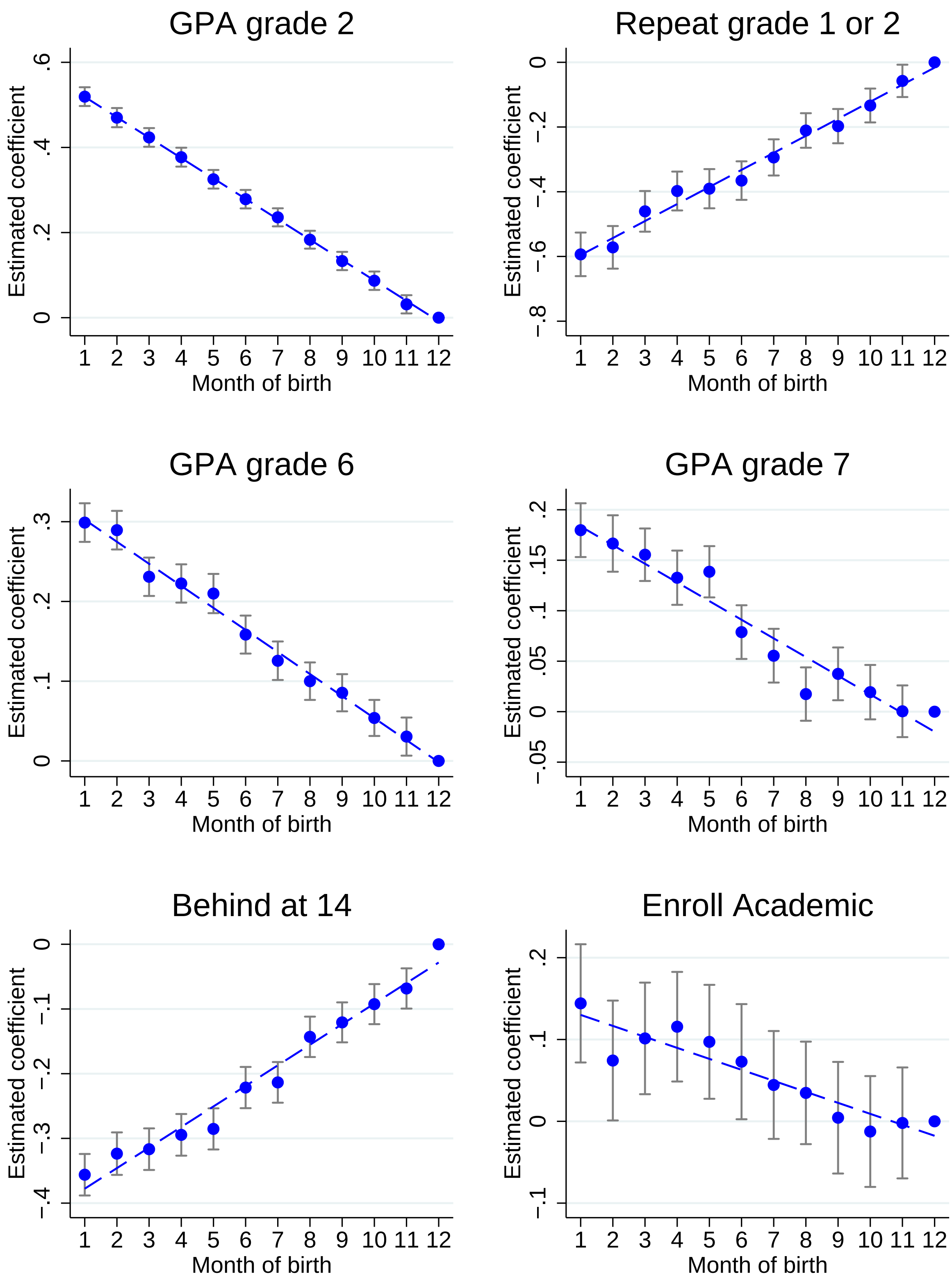
$$Y_i = aOA_i + X_i b + u_i, \qquad cor(u_i, OA_i) \neq 0 \tag{3.8}$$

$$OA_i = \gamma A_i + X_i \delta + v_i, \qquad cor(v_i, A_i) = 0, \quad cor(u_i, A_i) = 0 \tag{3.9}$$

The two stage least squares estimation estimates first $\widehat{\gamma}$ and $\widehat{\delta}$ in equation (3.9) and then replaces OA_i in equation (3.8) with $\widehat{OA_i} = \widehat{\gamma} A_i + X_i \widehat{\delta}$. Given that $\widehat{A_i}$ is uncorrelated with the error $(u_i + v_i)$, this makes possible a consistent estimation of a.

The reduced form approach consists in replacing age in equation (3.8) with the RHS

Figure 3.6: Estimated effect of month of birth on educational outcomes



The figures plot the estimated coefficients for month of birth dummies and their confidence interval. December is the baseline category.

of equation (3.9):

$$Y_i = a\gamma A_i + X(b + a\delta) + \alpha v_i + u \tag{3.10}$$

$$= \alpha A_i + X\beta + \epsilon_i, \qquad cor(\epsilon_i, A_i) = 0 \tag{3.11}$$

$\alpha = a\gamma$ can be consistently estimated, but in general it is different from a if $\gamma \neq 1$. In our case however γ is really close to 1, because $OA_i = A_i$ for 99% of students. Therefore we can expect $\widehat{\alpha}$ to deliver a good approximation of a, even if slightly smaller. The intuition is confirmed by results in Table 3.10. The first two columns of the table compare estimated coefficients for the regression of evaluations in grade 2 on age at entry and covariates. Column (RF) replicates the first column of Table 3.6 using the subsample of students born from 2003 to 2005. In column (2SLS), the observed age at enrollment OE_i is instrumented with the expected age A_i.

The estimated effect of age at entry is almost the same using the two approaches: it is 0.57 with the reduced form estimation and 0.58 with the two stage using instrumental variable approach. The third and fourth columns compare results for grade repetition in the first two grades of primary education, using the subsample of students born in 2003 and 2004. Again, the estimated effect is almost the same (-4 p.p. and -4.1 p.p. respectively). To conclude, we can regard the estimations obtained by reduced form as a lower bound of the true effect.

Appendix 3.B Private schools

Some of the analyses discussed in Section 3.5 can be replicated exploiting also data for students enrolled in private schools. Results are shown in Table 3.11. The regressions in columns (1) are performed using both students enrolled in public and private schools, while those in columns (2) are restricted to students in private schools.

Being one year older increases external evaluations in grade 6 of students enrolled in private schools by 0.33 standard deviations. The estimated coefficient is quite similar to the one found for public schools. In grade 10 the effect is significant and sizeable (0.11 standard deviations), but somewhat smaller than what found for public schools. However, it is important to recall that only students who do not dropout undertake evaluations in grade 10, therefore some selection bias may affect the estimate.

The age at entry has a large effect also on the probability of being behind at 14 (the average marginal effect is -8.5 p.p.) and enrolling in academic upper secondary education (the average marginal effect is 4.3 p.p.). Those figures are comparable with the finding for public schools. It is not surprising that they are slightly smaller because, as shown in column (1), students enrolled in private schools are less likely to be retained and more

Table 3.10: Evaluations and grade repetition. Instrumental variable approach.

	Evaluations (RF)	Evaluations (2SLS)	Repeat grade 1 or 2 (RF)	Repeat grade 1 or 2 (2SLS)	First stage (2SLS)
Age at entry (expected)	0.570		-0.040		0.988
	(0.009)**		(0.003)**		(0.001)**
Age at entry (observed)		0.577		-0.041	
		(0.010)**		(0.003)**	
Female	0.161	0.163	-0.011	-0.011	-0.003
	(0.006)**	(0.006)**	(0.001)**	(0.001)**	(0.000)**
Immigrant	-0.162	-0.165	0.027	0.027	0.005
	(0.013)**	(0.013)**	(0.003)**	(0.003)**	(0.001)**
Parents' edu avg	0.330	0.331	-0.028	-0.028	-0.002
	(0.008)**	(0.008)**	(0.002)**	(0.002)**	(0.001)**
Parents' edu high	0.628	0.630	-0.039	-0.039	-0.004
	(0.009)**	(0.009)**	(0.002)**	(0.002)**	(0.001)**
N	111,548	111,548	65,987	65,987	111,548

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. In columns (2SLS) the observed enrollment age OA_i is instrumented with the expected age A_i.

likely to undertake academic education.

Appendix 3.C Additional tables and figures

Table 3.11: Evaluations and grade repetition. Private schools.

	Evaluations gr. 6		Evaluations gr. 10		Behind at 14		Enroll Academic	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Age at entry	0.339	0.329	0.174	0.114	-0.374	-0.394	0.153	0.142
	(0.007)**	(0.011)**	(0.015)**	(0.020)**	(0.009)**	(0.017)**	(0.021)**	(0.039)**
					[-0.099]	[-0.085]	[0.053]	[0.043]
					<0.002>	<0.004>	<0.007>	<0.012>
Public school	-0.240		-0.352		0.184		-0.389	
	(0.013)**		(0.019)**		(0.006)**		(0.024)**	
					[0.049]		[-0.134]	
					<0.002>		<0.008>	
Parents' edu avg	0.459	0.386	0.446	0.357	-0.450	-0.442	0.548	0.556
	(0.008)**	(0.015)**	(0.014)**	(0.021)**	(0.006)**	(0.011)**	(0.015)**	(0.030)**
					[-0.119]	[-0.095]	[0.189]	[0.169]
					<0.002>	<0.002>	<0.005>	<0.009>
Parents' edu high	0.840	0.711	0.873	0.707	-0.963	-0.912	1.167	1.130
	(0.011)**	(0.019)**	(0.018)**	(0.024)**	(0.008)**	(0.012)**	(0.021)**	(0.037)**
					[-0.255]	[-0.196]	[0.402]	[0.344]
					<0.002>	<0.003>	<0.006>	<0.010>
N	190,371	73,287	42,816	18,137	292,500	112,421	48,959	15,490

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. In columns (1), both students enrolled in public and private schools are included in the analysis. In columns (2), only students enrolled in private schools are included.

Table 3.12: Evaluations by subjects

	Mathematics			Catalan			Spanish			English		
	2	6	7	2	6	7	2	6	7	2	6	7
Age at entry	0.536	0.299	0.201	0.504	0.308	0.198	0.529	0.311	0.206	0.489	0.284	0.195
	(0.008)**	(0.009)**	(0.010)**	(0.008)**	(0.009)**	(0.010)**	(0.009)**	(0.009)**	(0.009)**	(0.008)**	(0.009)**	(0.009)**
Female	-0.082	0.066	0.157	0.229	0.336	0.365	0.249	0.353	0.364	0.230	0.325	0.354
	(0.005)**	(0.006)**	(0.007)**	(0.005)**	(0.005)**	(0.007)**	(0.005)**	(0.005)**	(0.007)**	(0.005)**	(0.006)**	(0.007)**
Immigrant	-0.158	-0.174	-0.201	-0.232	-0.257	-0.243	-0.275	-0.265	-0.262	-0.091	-0.151	-0.191
	(0.011)**	(0.011)**	(0.013)**	(0.011)**	(0.011)**	(0.014)**	(0.011)**	(0.011)**	(0.013)**	(0.011)**	(0.011)**	(0.014)**
Parents' edu avg	0.297	0.312	0.297	0.301	0.312	0.283	0.285	0.313	0.290	0.292	0.319	0.309
	(0.006)**	(0.007)**	(0.008)**	(0.006)**	(0.007)**	(0.008)**	(0.006)**	(0.007)**	(0.008)**	(0.006)**	(0.007)**	(0.008)**
Parents' edu high	0.571	0.633	0.636	0.590	0.633	0.615	0.486	0.589	0.588	0.573	0.635	0.642
	(0.008)**	(0.009)**	(0.012)**	(0.008)**	(0.009)**	(0.012)**	(0.008)**	(0.009)**	(0.012)**	(0.008)**	(0.009)**	(0.011)**
R^2	0.09	0.09	0.09	0.11	0.12	0.12	0.10	0.12	0.12	0.10	0.11	0.12
N	162,048	137,300	108,389	161,910	137,059	108,349	161,874	137,162	108,153	160,782	137,023	106,938

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. Cohort FE included. Standard errors clustered by school. For each column the sample used is the same than in Table 3.6.

Table 3.13: Evaluations in primary education. Peer characteristics included as regressors

	Grade 2	Grade 4	Grade 6 (int.)	Grade 6 (ext.)
Age at entry	0.575	0.409	0.332	0.349
	(0.008)**	(0.008)**	(0.009)**	(0.009)**
Female	0.168	0.218	0.291	0.135
	(0.005)**	(0.005)**	(0.005)**	(0.006)**
Immigrant	-0.308	-0.322	-0.328	-0.333
	(0.010)**	(0.010)**	(0.010)**	(0.011)**
Parents' edu avg	0.383	0.403	0.392	0.396
	(0.006)**	(0.006)**	(0.006)**	(0.009)**
Parents' edu high	0.747	0.799	0.807	0.736
	(0.007)**	(0.007)**	(0.008)**	(0.009)**
Mean age at entry	-0.523	-0.335	-0.270	0.183
	(0.025)**	(0.028)**	(0.030)**	(0.071)**
% females	-0.158	-0.205	-0.235	0.141
	(0.014)**	(0.016)**	(0.018)**	(0.050)**
% immigrants	0.469	0.466	0.429	-0.322
	(0.015)**	(0.014)**	(0.014)**	(0.048)**
% parents' edu high	-0.629	-0.667	-0.693	0.571
	(0.011)**	(0.011)**	(0.013)**	(0.042)**
R^2	0.15	0.15	0.15	0.17
N	160,438	148,472	136,281	117,752

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. Cohort FE included. Standard errors clustered by school. Peer characteristics are computed at the class level. For each column the sample used is the same than in Table 3.6.

Table 3.14: Evaluations in primary and lower secondary education. First-time evaluations are used for repeaters.

	grade 2	grade 4	grade 6		grade 7	grade 8	grade 9	grade 10	
			(int.)	(ext.)				(int.)	(ext.)
Age (0-1)	0.593	0.415	0.333	0.353	0.229	0.182	0.161	0.154	0.227
	(0.008)**	(0.008)**	(0.009)**	(0.010)**	(0.009)**	(0.010)**	(0.011)**	(0.012)**	(0.021)**
Female	0.172	0.218	0.293	0.135	0.347	0.337	0.341	0.333	0.076
	(0.005)**	(0.005)**	(0.005)**	(0.006)**	(0.007)**	(0.007)**	(0.008)**	(0.008)**	(0.012)**
Immigrant	-0.218	-0.226	-0.231	-0.411	-0.251	-0.231	-0.250	-0.332	-0.735
	(0.012)**	(0.011)**	(0.011)**	(0.014)**	(0.014)**	(0.014)**	(0.013)**	(0.013)**	(0.025)**
Parents' edu avg	0.337	0.354	0.345	0.436	0.327	0.309	0.261	0.196	0.402
	(0.007)**	(0.007)**	(0.007)**	(0.010)**	(0.008)**	(0.009)**	(0.008)**	(0.008)**	(0.016)**
Parents' edu high	0.640	0.680	0.684	0.840	0.692	0.670	0.598	0.515	0.861
	(0.008)**	(0.008)**	(0.009)**	(0.012)**	(0.012)**	(0.012)**	(0.011)**	(0.012)**	(0.020)**
R^2	0.12	0.12	0.12	0.15	0.14	0.12	0.11	0.10	0.20
N	160,369	148,426	136,231	117,566	106,567	102,870	96,068	85,704	24,367

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. Cohort FE included. Standard errors clustered by school. For each column the sample used is the same than in Table 3.6.

Table 3.15: Quantile regressions

	grade 2	grade 4	grade 6		grade 7	grade 8	grade 9	grade 10	
			(int.)	(ext.)				(int.)	(ext.)
q15	0.480	0.321	0.230	0.421	0.136	0.119	0.100	0.078	0.206
	(0.012)**	(0.011)**	(0.013)**	(0.021)**	(0.014)**	(0.018)**	(0.022)**	(0.018)**	(0.036)**
q25	0.570	0.376	0.280	0.424	0.160	0.111	0.106	0.098	0.217
	(0.009)**	(0.012)**	(0.012)**	(0.016)**	(0.012)**	(0.016)**	(0.016)**	(0.013)**	(0.029)**
q35	0.612	0.429	0.334	0.406	0.178	0.132	0.105	0.083	0.179
	(0.008)**	(0.012)**	(0.012)**	(0.013)**	(0.011)**	(0.011)**	(0.013)**	(0.010)**	(0.028)**
q50	0.678	0.475	0.380	0.382	0.216	0.162	0.129	0.121	0.206
	(0.009)**	(0.011)**	(0.010)**	(0.010)**	(0.011)**	(0.009)**	(0.009)**	(0.008)**	(0.029)**
q65	0.692	0.489	0.415	0.343	0.261	0.196	0.161	0.141	0.200
	(0.010)**	(0.014)**	(0.011)**	(0.008)**	(0.013)**	(0.009)**	(0.013)**	(0.010)**	(0.022)**
q75	0.657	0.479	0.405	0.302	0.291	0.238	0.205	0.172	0.169
	(0.009)**	(0.015)**	(0.012)**	(0.009)**	(0.013)**	(0.012)**	(0.013)**	(0.014)**	(0.025)**
q85	0.579	0.422	0.376	0.262	0.303	0.261	0.231	0.206	0.148
	(0.010)**	(0.013)**	(0.012)**	(0.011)**	(0.015)**	(0.015)**	(0.015)**	(0.015)**	(0.022)**
N	160,438	148,472	136,281	117,752	106,707	102,960	96,170	86,025	32,053

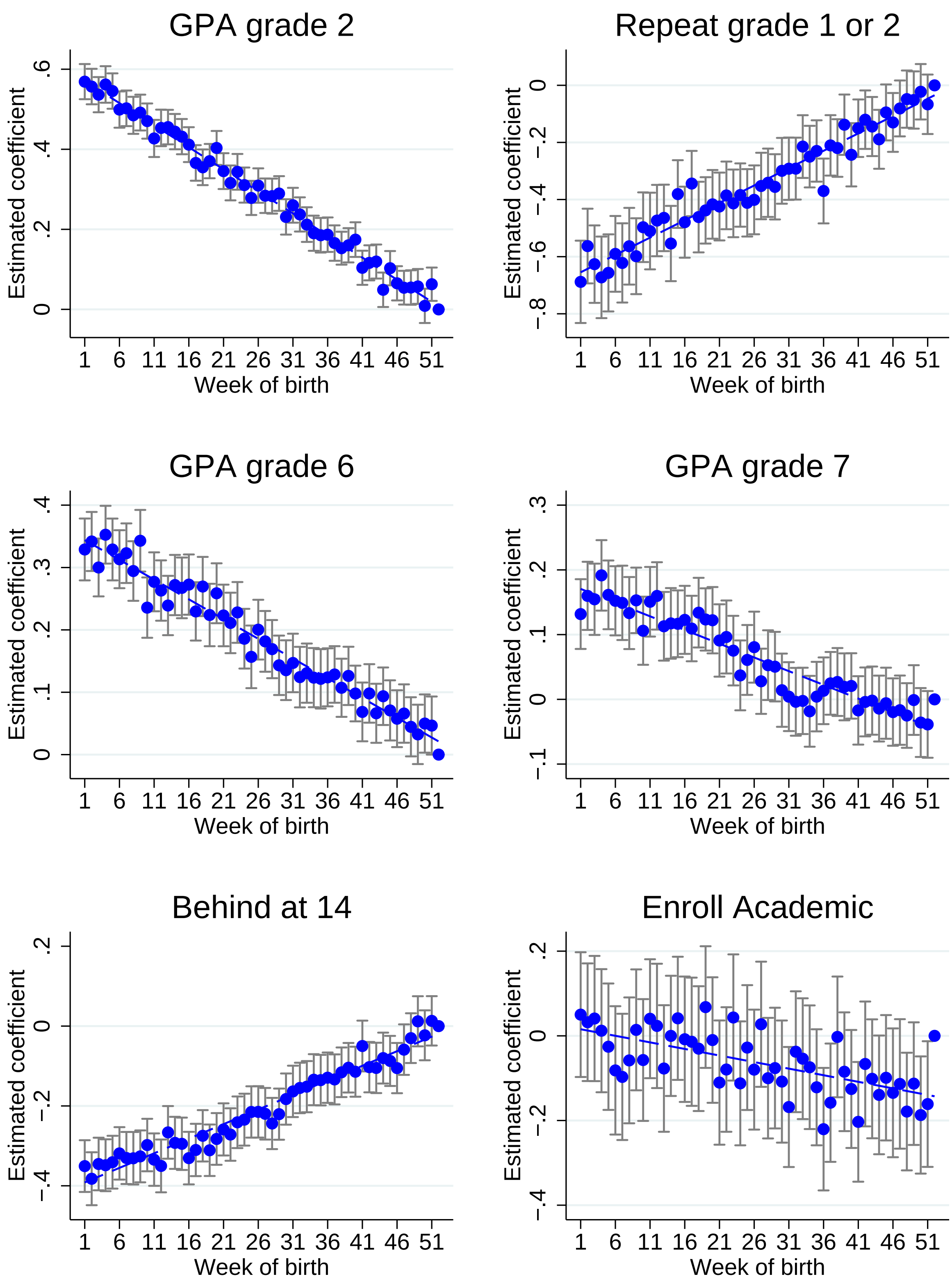
Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$. Regressors include dummies for gender, immigrant, parental education, and cohort. For each column the sample used is the same than in Table 3.6.

Table 3.16: Regressions with month or week dummies

	GPA grade 2		GPA grade 6		GPA grade 7		Repeat grade 1 or 2		Behind at 14		Enroll Academi	
	(M)	(W)	(M)	(W)	(M)	(W)	(M)	(W)	(M)	(W)	(M)	(W)
January / Week 1	0.519	0.569	0.299	0.329	0.180	0.132	-0.593	-0.688	-0.356	-0.351	0.144	0.050
	(0.011)**	(0.022)**	(0.012)**	(0.025)**	(0.014)**	(0.028)**	(0.034)**	(0.074)**	(0.016)**	(0.033)**	(0.037)**	(0.075
November / Week 51	0.031	0.063	0.031	0.047	0.000	-0.039	-0.057	-0.066	-0.068	0.013	-0.002	-0.16
	(0.011)**	(0.021)**	(0.012)*	(0.024)+	(0.013)	(0.026)	(0.025)*	(0.053)	(0.016)**	(0.032)	(0.035)	(0.076)
Female	0.170	0.170	0.293	0.293	0.339	0.339	-0.121	-0.121	-0.283	-0.283	0.416	0.416
	(0.005)**	(0.005)**	(0.005)**	(0.005)**	(0.007)**	(0.007)**	(0.013)**	(0.013)**	(0.007)**	(0.007)**	(0.016)**	(0.016)
Immigrant	-0.203	-0.204	-0.227	-0.227	-0.243	-0.243	0.322	0.322	0.523	0.523	-0.554	-0.55
	(0.011)**	(0.011)**	(0.011)**	(0.011)**	(0.014)**	(0.014)**	(0.020)**	(0.020)**	(0.008)**	(0.008)**	(0.024)**	(0.024)
Parents' edu avg	0.326	0.326	0.342	0.341	0.323	0.323	-0.364	-0.365	-0.396	-0.397	0.503	0.503
	(0.006)**	(0.006)**	(0.007)**	(0.007)**	(0.008)**	(0.008)**	(0.016)**	(0.016)**	(0.007)**	(0.007)**	(0.017)**	(0.017)
Parents' edu high	0.621	0.621	0.679	0.679	0.680	0.680	-0.705	-0.705	-0.899	-0.899	1.110	1.112
	(0.008)**	(0.008)**	(0.009)**	(0.009)**	(0.012)**	(0.012)**	(0.024)**	(0.024)**	(0.011)**	(0.011)**	(0.025)**	(0.025)
N	160,438	160,438	136,281	136,281	106,707	106,707	163,930	163,930	178,871	178,871	33,624	33,62

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

Figure 3.7: Estimated effect of week of birth on educational outcomes



The figures plot the estimated coefficients for week of birth dummies and their confidence interval. Week 52 is the baseline category.

Table 3.17: Regressions using students born around the cutoff date

	GPA 2	GPA 6	GPA 7	Repeat 1-2	Behind 14	Enroll Academic
Old	0.531	0.312	0.168	-0.593	-0.370	0.121
	(0.016)**	(0.017)**	(0.019)**	(0.049)**	(0.023)**	(0.050)*
				[-0.043]	[-0.108]	[0.041]
Female	0.172	0.305	0.327	-0.157	-0.273	0.422
	(0.016)**	(0.017)**	(0.020)**	(0.042)**	(0.023)**	(0.048)**
				[-0.011]	[-0.080]	[0.145]
Immigrant	-0.181	-0.257	-0.302	0.326	0.497	-0.581
	(0.027)**	(0.024)**	(0.028)**	(0.057)**	(0.028)**	(0.071)**
				[0.024]	[0.145]	[-0.200]
Parents' edu avg	0.344	0.343	0.323	-0.387	-0.393	0.453
	(0.019)**	(0.020)**	(0.022)**	(0.053)**	(0.026)**	(0.056)**
				[-0.028]	[-0.115]	[0.156]
Parents' edu high	0.640	0.665	0.633	-0.689	-0.848	1.129
	(0.021)**	(0.024)**	(0.029)**	(0.071)**	(0.036)**	(0.076)**
				[-0.050]	[-0.247]	[0.388]
N	13,300	11,543	8,856	13,572	14,978	2,694

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

Table 3.18: Regressions with interaction terms (I)

	Retention		Evaluations					
	repeat 1-2	behind 14	2	4	6	6-ext.	7	8
Age at entry	-0.671	-0.381	0.543	0.406	0.349	0.383	0.207	0.148
	(0.043)**	(0.021)**	(0.017)**	(0.017)**	(0.017)**	(0.022)**	(0.019)**	(0.019)**
Age X parents' edu avg	-0.068	-0.034	0.053	0.056	0.022	-0.012	-0.004	0.023
	(0.051)	(0.026)	(0.019)**	(0.019)**	(0.019)	(0.025)	(0.023)	(0.022)
Age X parents' edu high	0.007	0.019	0.027	-0.023	-0.004	-0.052	0.068	0.076
	(0.078)	(0.038)	(0.020)	(0.022)	(0.023)	(0.025)*	(0.025)**	(0.028)**
Age X female	0.085	0.018	0.028	0.015	0.009	0.013	0.044	0.051
	(0.048)+	(0.024)	(0.016)+	(0.016)	(0.017)	(0.020)	(0.020)*	(0.019)**
Age X immigrant	0.065	0.002	-0.075	-0.137	-0.165	-0.133	-0.122	-0.115
	(0.057)	(0.028)	(0.025)**	(0.023)**	(0.023)**	(0.030)**	(0.026)**	(0.025)**
Parents' edu avg	-0.337	-0.380	0.300	0.321	0.331	0.441	0.326	0.294
	(0.027)**	(0.014)**	(0.011)**	(0.011)**	(0.012)**	(0.016)**	(0.014)**	(0.014)**
Parents' edu high	-0.707	-0.907	0.608	0.682	0.681	0.863	0.647	0.623
	(0.038)**	(0.021)**	(0.013)**	(0.014)**	(0.014)**	(0.018)**	(0.017)**	(0.019)**
Female	-0.158	-0.291	0.156	0.211	0.288	0.128	0.317	0.313
	(0.023)**	(0.013)**	(0.009)**	(0.009)**	(0.010)**	(0.012)**	(0.012)**	(0.011)**
Immigrant	0.299	0.522	-0.167	-0.153	-0.145	-0.347	-0.182	-0.167
	(0.031)**	(0.016)**	(0.017)**	(0.015)**	(0.016)**	(0.021)**	(0.019)**	(0.018)**
N	163,910	178,871	160,438	148,472	136,281	117,752	106,707	102,960

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

Table 3.19: Regressions with interaction terms (II)

	Graduate	Enroll		Special Needs	
		Academic	Vocational	gr 1-2	gr 7-8
Age at entry	0.117	0.232	-0.039	-0.216	-0.230
	(0.048)*	(0.047)**	(0.049)	(0.042)**	(0.043)**
Age X parents' edu avg	0.113	0.138	-0.146	-0.120	0.043
	(0.058)+	(0.055)*	(0.060)*	(0.056)*	(0.060)
Age X parents' edu high	-0.055	-0.037	-0.092	0.136	0.111
	(0.077)	(0.074)	(0.080)	(0.066)*	(0.089)
Age X female	-0.134	-0.165	-0.074	-0.022	-0.037
	(0.054)*	(0.051)**	(0.055)	(0.050)	(0.051)
Age X immigrant	-0.074	-0.178	0.143	-0.100	0.012
	(0.060)	(0.066)**	(0.063)*	(0.070)	(0.060)
Parents' edu avg	0.308	0.435	-0.085	-0.159	-0.381
	(0.032)**	(0.033)**	(0.036)*	(0.034)**	(0.035)**
Parents' edu high	0.774	1.128	-0.501	-0.455	-0.641
	(0.050)**	(0.047)**	(0.047)**	(0.041)**	(0.051)**
Female	0.416	0.498	-0.211	-0.324	-0.271
	(0.031)**	(0.031)**	(0.031)**	(0.027)**	(0.029)**
Immigrant	-0.521	-0.466	-0.216	0.051	0.064
	(0.038)**	(0.039)**	(0.038)**	(0.041)	(0.037)+
N	33,624	33,624	33,624	149,142	115,204

Notes. + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$.

Bibliography

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.

Ahn, T., P. Arcidiacono, A. Hopson, and J. Thomas (2016, June). Equilibrium grade inflation with implications for female interest in stem majors. Mimeo.

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

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

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

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

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

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

Angrist, J. D. and A. B. Keueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

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.

Arellano, M. and M. Weidner (2016). Instrumental Variable Quantile Regressions in Large Panels with Fixed Effects. Work in progress.

Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar (2009). Large stakes and big mistakes. *The Review of Economic Studies 76*(2), 451.

Attali, Y., Z. Neeman, and A. Schlosser (2011, May). Rise to the Challenge or Not Give a Damn: Differential Performance in High vs. Low Stakes Tests. IZA Discussion Papers 5693, Institute for the Study of Labor (IZA).

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.

Azmat, G., C. Calsamiglia, and N. Iriberri (2016). Gender differences in response to big stakes. *Journal of the European Economic Association 14*(6), 1372–1400.

Azmat, G. and N. Iriberri (2010, August). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics 94*(7-8), 435–452.

Bedard, K. and E. Dhuey (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics 121*(4), 1437–1472.

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

Berniell, I. and R. Estrada (2017, November). Poor Little Children: The Socioeconomic Gap in Parental Responses to School Disadvantage. Caf working papers.

Betts, J. R. and J. Grogger (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review 22*(4), 343 – 352.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2011, May). Too Young to Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics 93*(2), 455–467.

Bobba, M. and V. Frisancho (2016). Learning About Oneself: The Effects of Signaling Academic Ability on School Choice. Mimeo.

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

Burke, M. A. and T. R. Sass (2013). Classroom Peer Effects and Student Achievement. *Journal of Labor Economics 31*(1), 51 – 82.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal 14*(4), 909–946.

Calsamiglia, C., C. Fu, and M. Güell (2014, October). Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives. Working Papers 811, Barcelona Graduate School of Economics.

Calsamiglia, C. and M. Güell (2014, July). The Illusion of School Choice: Empirical Evidence from Barcelona. Working Papers 810, Barcelona Graduate School of Economics.

Calsamiglia, C. and A. Loviglio (2018, May). Grading on a Curve: When Having Good Peers Is Not Good. Mimeo.

Calsamiglia, C. and A. Loviglio (2019, May). Maturity and School Outcomes in an Inflexible System: Evidence from Catalonia. Mimeo.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013, 05). From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica 81*(3), 855–882.

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

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

Cook, P. J. and S. Kang (2016, January). Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation. *American Economic Journal: Applied Economics 8*(1), 33–57.

Crawford, C., L. Dearden, and C. Meghir (2010, May). When you are born matters: the impact of date of birth on educational outcomes in England. IFS Working Papers W10/06, Institute for Fiscal Studies.

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.

Diamond, R. and P. Persson (2016, April). The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests. NBER Working Papers 22207, National Bureau of Economic Research, Inc.

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.

Ebenstein, A., V. Lavy, and S. Roth (2016, October). The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics 8*(4), 36–65.

Elder, T. E. and D. H. Lubotsky (2009). Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers. *Journal of Human Resources 44*(3).

Elsner, B. and I. E. Isphording (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics 35*(3), 787–828.

Epple, D. and R. Romano (2011). Peer effects in education: A survey of the theory and evidence. *Handbook of social economics 1*(11), 1053–1163.

Estevan, F., T. Gall, P. Legros, and A. F. Newman (2014, November). College Admission and High School Integration. Working Papers, Department of Economics 2014 - 26, University of Sao Paulo (FEA-USP).

Feld, J. and U. Zölitz (2016). Understanding peer effects - On the nature, estimation and channels of peer effects. Research Memorandum 002, Maastricht University, Graduate School of Business and Economics (GSBE).

Figlio, D. and M. Lucas (2004, 8). Do high grading standards affect student performance? *Journal of Public Economics 88*(9-10), 1815–1834.

Frandsen, B. R. (2017). *Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete*, Chapter 7, pp. 281–315.

Fredriksson, P. and B. Öckert (2014). Lifecycle Effects of Age at School Start. *Economic Journal 124*(579), 977–1004.

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

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

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

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

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

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

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

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

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

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

Kinsler, J., R. Pavan, and R. DiSalvo (2014, December). Distorted beliefs and parental investment in children. Mimeo.

Lavy, V. (2008, October). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics 92*(10-11), 2083–2105.

Lavy, V. and E. Sand (2015, January). On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers Stereotypical Biases. NBER Working Papers 20909, National Bureau of Economic Research, Inc.

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

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Mayer, S. E. and C. Jencks (1989). Growing up in poor neighborhoods: How much does it matter? *Science 243*(4897), 1441–1445.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698 – 714. The regression discontinuity design: Theory and applications.

McEwan, P. J. and J. S. Shapiro (2008). The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. *Journal of Human Resources 43*(1).

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

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

Ponzo, M. and V. Scoppa (2014). The long-lasting effects of school entry age: Evidence from Italian students. *Journal of Policy Modeling 36*(3), 578–599.

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

Puhani, P. and A. Weber (2007, May). Does the early bird catch the worm? *Empirical Economics 32*(2), 359–386.

Rangvid, B. S. (2015). Systematic differences across evaluation schemes and educational choice. *Economics of Education Review 48*, 41 – 55.

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

Sacerdote, B. (2011). *Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?*, Volume 3 of *Handbook of the Economics of Education*, Chapter 4, pp. 249–277. Elsevier.

Schneeweis, N. and M. Zweimüller (2014). Early tracking and the misfortune of being young. *The Scandinavian Journal of Economics 116*(2), 394–428.

Tincani, M. (2015, June). Heterogeneous peer effects and rank concerns: Theory and evidence. Mimeo.

Tran, A. and R. Zeckhauser (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics 96*(9-10), 645–650.

Weinhardt, F. and R. Murphy (2016, February). Top of the Class: The Importance of Ordinal Rank. Mimeo.

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.