

Parental environment and student achievement: Does a Matthew effect exist? *

Gaëlle Aymeric,[†] Emmanuelle Lavaine[‡] and Brice Magdalou[§]

October 22, 2025

Abstract

This paper investigates the causal impact of the parental environment on the student's academic performance in mathematics, literature and English (as a foreign language), using a new database covering all children aged 8 to 15 of the Madrid community, from 2016 to 2019. Parental environment refers here to the parents' level of education (i.e. the skills they acquired before bringing up their children), and parental investment (the effort made by parents to bring up their children). We distinguish the *persistent effect* of the parental environment from the so-called *Matthew effect*, which describes a possible tendency for the impact of the parental environment to increase as the child grows up. Whatever the subject (mathematics, literature or English), our results are in line with most studies concerning the persistent effect: a favourable parental environment goes hand in hand with better results for the children. As regards the Matthew effect, the results differ between subjects: while the impact of the parental environment tends to diminish from the age of 8 to 15 in mathematics, it forms a bell curve in literature (first increasing, then decreasing) and increases steadily in English. This result, which is encouraging for mathematics and even literature, confirms the social dimension involved in learning a foreign language compared to more academic subjects.

JEL Classification Numbers: I24, D63. *Keywords:* Academic results; Parental environment; Equality of opportunity; Matthew effect.

* This paper is part of the research project *MaDimIn* (Contract ANR-24-CE26-3823-02), from which financial support is acknowledged. This paper was written when Brice Magdalou was on a temporary CNRS research assignment at the Aix-Marseille School of Economics. We acknowledge financial support from the French government under the “France 2030” investment plan managed by the French National Research Agency Grant ANR-17-EURE-0020, and by the Excellence Initiative of Aix-Marseille University - A*MIDEX. We are grateful to Elena Bárcena-Márín, Nicolas Gravel, Markus Jäntti, Jo Thori Lind, Vito Peragine, Rafael Salas, Daniel Santín and Alain Trannoy for their useful comments.

[†] Universidad Complutense de Madrid
E-mail: gaymeric@uclm.es.

[‡] CEE-M, Univ. Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France.
E-mail: emmanuelle.lavaine@umontpellier.fr.

[§] CNRS, AMSE, Marseille, France.
E-mail: brice.magdalou@umontpellier.fr.

‘To all those who have, more will be given, and they will have an abundance; but from those who have nothing, even what they have will be taken away’

[Matthew (the Bible 13:12, NRSV)]

1. Introduction

It is now well-documented that the cognitive and non-cognitive abilities developed in the early childhood drive the educational, social and professional success of people throughout their entire life. It is also well-recognized that the social background and the investment of parents in their own children impact ability acquisition, which partly explains the inequalities in academic performance across children. According to the Matthew effect,¹ in many spheres of life, ‘the rich get richer and the poor get poorer’. In the education field, this effect describes a possible tendency of initial advantages, in early life, to accumulate through time. Whereas the *persistent effect* of the parental environment on student achievement is now admitted, the theoretical and empirical literatures are more balanced on the existence of a possible *Matthew effect*. This paper aims at contributing to this debate.

Different channels can explain the impact of parental environment on children academic performance. The first is a possible *intergenerational transmission of cognitive skills*, which implies that the association between parents and children abilities can be partly driven by genetic.² A second is through the *parent’s level of education*. The children can benefit from the knowledge and diplomas acquired by their parents, but also from the related positive spillovers.³ A third channel is *parental investment*, which can be a major input in the child production of skills. In that case, a distinction has to be drawn between cognitive and non-cognitive skills. Cunha *et al.* (2010) found that the productivity of parental investment for cognitive skills is high in the early stages of education (before 6), but tends to significantly decrease after. At the opposite, the productivity of parental investment on non-cognitive skills is found to be higher at later stages. Finally, these channels can be exacerbated by a possible *assortative mating* of the parents.⁴

¹ The Matthew effect, a standard concept in sociology, has been popularized by Merton (1968). Rigney (2010) proposes a review of applications in several (social) sciences, including education.

² Hanushek *et al.* (2021) identify a causal connection between cognitive skills of the parents and their children, based on a Dutch survey on math and language skills. Sacerdote (2007) uses Korean American adoptees data to show that genetic factors explain 44% of the variation in educational attainment and 33% of the variation in income.

³ The survey proposed by Holmlund *et al.* (2011) concludes that the estimates of the causal effect of parent’s schooling on child’s schooling differ across studies, but also that selection is the main component of the inter-generational association. At the opposite, by using original Finnish data, Suhonen and Karhunen (2019) find a strong positive causal effect (of around 0.5) from parent’s to child’s attained years of education.

⁴ Bingley *et al.* (2022) find that 75% of the correlation in education attainment between parents and their children

The literature on the technology of skills formation, initiated by [Cunha and Heckman \(2007\)](#) and [Cunha *et al.* \(2010\)](#), tends to support the hypothesis of cumulative advantages. The authors propose and estimate a model where, at each stage of the child development, the inputs and the production technology can differ. They find that *self-productivity* (the stock of skills produced at one stage augment the skills attained at later stages), for both cognitive and non-cognitive skills, becomes stronger as the child becomes older. They also observe *dynamic complementary* (the productivity of an investment can be raised by skills produced at previous stages), but with a decrease in substitutability between investment in one period and the existing stock of skills. Hence, it is more and more difficult to compensate for initial endowment differences, which can imply an increasing attainment gap between advantaged and disadvantaged children.⁵

On the other hand, the equality of opportunity literature suggests that, in moving towards adulthood, the child is able to free herself (at least partially) from some external factors that have determined her previous achievements. In the same vein of the age of sexual consent or the age of criminal responsibility, this theory refers to what we can call an *age (of consent) for responsible choices* ([Roemer and Trannoy, 2016](#); [Hufe *et al.*, 2017](#)). In early childhood, the child cannot be held responsible for her behaviors and achievements as they result from *circumstances* not under her control.⁶ In contrast, one can assume that an adult is able to set out personal objectives and to take free and enlightened decisions (whatever her background), such as the level of effort she decides to put at work. This prerequisite is actually necessary for the existence of freedom in itself, by considering that the life trajectory is not fully deterministic. Of course the age of consent is a normative concept here, not a precise age threshold, and it is debatable to fix it before adulthood.⁷ But with this concept in hands, one can hypothesize that initial disadvantages or the parental influence can be partially mitigated, throughout schooling, by the emancipation of the child as she grows up.⁸

is driven by the joint contribution of the parents (as compared to the contribution of each parent independently). [Eika *et al.* \(2019\)](#) also find that educational assortative mating has declined among college graduates in the US since the 1960s, while it has progressively increased among the low-educated (a trend also true for the other countries studied: Denmark, Germany, the United Kingdom, and Norway).

⁵ One main policy recommendation resulting from these estimations is that successful adolescent remediation strategies for disadvantaged children should focus on fostering non-cognitive skills.

⁶ Laziness at school, for instance, might be explained by a home environment which is neither stimulating nor rewarding.

⁷ [Roemer and Trannoy \(2016\)](#) emphasize that it is controversial to use years of education as an effort variable (hence after the age of consent) until the end of secondary education, and consider that only tertiary education is immune to this criticism. [Hufe *et al.* \(2017\)](#) fix this age between 12 and 16, and recalculate the fraction of income inequality due to circumstances in the US and the UK, by considering that all the childhood achievements before the age of consent is a circumstance.

⁸ By studying the academic performance of students who are the first in their family to attend university, [Edwards *et al.* \(2022\)](#) establish that some non-cognitive skills such as conscientiousness or extraversion, predict academic performance almost as strongly as standardised university admissions test scores. One can assume that such

In this paper, we investigate a rich and never exploited database on the Madrid Community (Spain) to analyse the impact of parental environment (including parental highest schooling and parental investment) on their child’s academic performance in three subjects (mathematics, literature and English as foreign language) and its change at three different education grades (Grades 3, 6 and 10, respectively about 8, 11 and 15 years old), over four academic years (from 2016 to 2019). We have combined data from various sources, provided by the Ministry of Education and Research of the Community of Madrid. First, for each grade, we have the scores of the students in each subject. The scores are normalised following a method comparable to the one used by the OECD’s Programme for International Student Assessment (PISA). Then, we have the data from two questionnaires, one sent to the students and another sent to the parents. We obtain various descriptive observations for the students and the parents (gender, country of birth, ...) as well as behavioral observations, which can be used as proxy of the parental investment and child’s effort. This repeated cross-sectional dataset covers the overall Madrid Community, including private and public schools, and gathering data for more than 320,000 students.

Our empirical results contribute to the literature in several main directions. *First*, through a linear regression with fixed effects, we observe that the parental environment—both parents’ highest level of education and their investment—is a strong predictor of the child’s score in mathematics, literature, and English. Hence, a favorable parental environment goes hand in hand with better results for the children, what we call the persistent effect. *Second*, by using interaction components to observe how this influence evolves as the child progresses through school, we find mixed but informative results on the Matthew effect: whereas the parental influence on a child’s score decreases from age 11 to 15 in mathematics and follows a bell curve in literature, it is continuously and significantly increasing in English. *Third*, to address potential endogeneity and establish causality, we employ an instrumental variable (IV) analysis using the historical gender gap in tertiary education in the parents’ country of origin as an instrument. These IV estimates reinforce our initial findings, providing stronger causal evidences for the persistent effect and the mixed picture regarding the Matthew effect. Finally, our results also show that while a child’s own effort (proxied by time spent on homework) becomes increasingly impactful on academic results as they get older in all subjects, this effect is slightly weaker in English. This result confirms the special status of second language learning in schools ([Gardner, 1968](#)): The acquisition of a new language is a highly social process, determined by the environment in which the child lives, and which can be a source of important inequality of opportunity. In contrast, remediation programs for adolescent could be effective in mathematics or literature ([Bahr, 2007, 2008](#)) by fostering, for instance, non-cognitive skills, which serve as catalysts for cognitive development, such as perseverance, conscientiousness, or curiosity. Such optimism is supported by the clear indication, obtained

skills do not necessarily result from parental investment.

from our data, that effort is more and more successful after some age (for all the subjects).

The rest of the paper is organized as follows. We present in Section 2 our data and some descriptive results. Section 3 investigates the impact of parental highest schooling on student achievement, on the basis of two linear regression strategies (with or without interaction components). In this section, we also analyze the impact of parental investment and child's effort. In Section 4, we continue our investigation into the impact of parental highest schooling, using an instrumental variable approach, to avoid any potential endogeneity issues. Finally, we discuss in Section 5 our main results, in the light of a related literature, mainly in psychology. We also present some implications in terms of educational policies.

2. Data

This paper is based on a rich database that has never been used in academic research. From 2016 to 2019, the Ministry of Education and Research of Madrid Community (Spain)⁹ has organized annual exams for all the students of the community in Grade 3 (8 years old), Grade 6 (11 years old) and Grade 10 (15 years old; not assessed in 2016). In parallel with these examinations, four questionnaires were organized for the various stakeholders: one addressed to the parents, one addressed to the student (Grade 6 and Grade 10), one addressed to the school director, and one addressed to the teachers. The main aim of these questionnaires was to assess people's own feelings about the quality of the educational system, but also to evaluate people's involvement (such as the time parents devote to their child's education, or the weekly time children spend on homework). Surprisingly enough, these data have never been used in academic research up to date.

The first contribution of the present paper is to gather a set of disparate files and documents. We obtain a unique and harmonized cross-sectional database, covering more than 320,000 students. This database has a number of advantages, over most existing databases. First, this is not a simple survey as they cover all pupils in the Madrid community (whether in private or public schools). Then, pupils have a common identifier on examinations and questionnaires, so that it is possible to combine quantitative data on academic performance with more qualitative data, describing the educational environment in some detail.

This study focuses on three subjects: mathematics, literature and English (foreign language). As with many of the world's leading education surveys,¹⁰ the final exam score in each subject is calculated on the basis of the *Item Response Theory* (IRT). That refers to a family of mathematical

⁹ Consejería de Educación y Investigación de la Comunidad de Madrid.

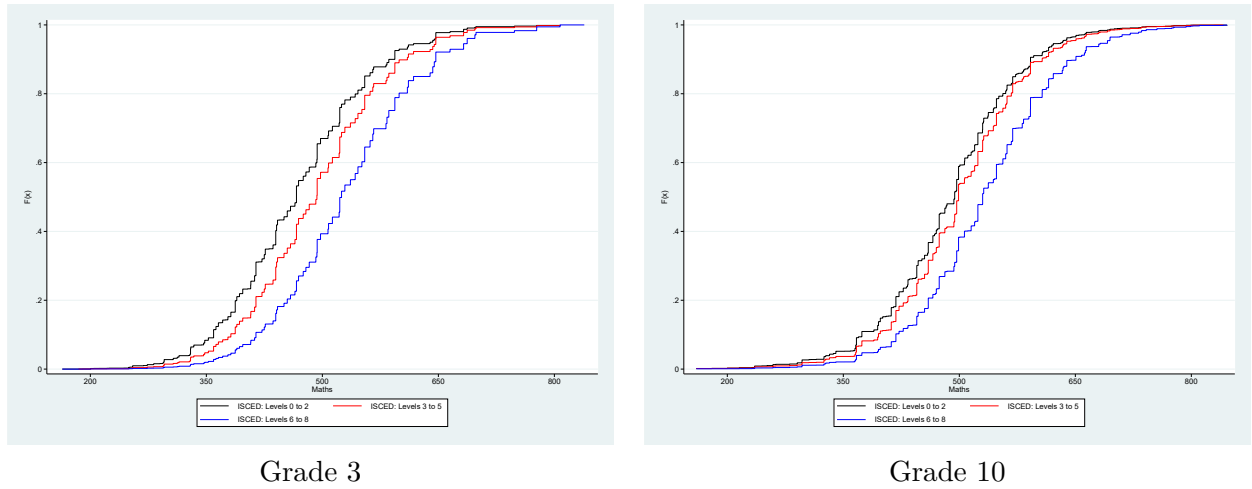
¹⁰ A few examples: PISA, TALIS or PIAAC, for the OECD; TIMSS or PIRLS for the International Association for the Evaluation of Educational Achievement (IEA); TOEFL and Cambridge Certification.

models that attempt to explain the relationship between a candidate’s response to an item and that candidate’s aptitude or skills. In this study, the *Partial Credit Model* (PCM) is implemented (see [Masters and Wright, 1997](#)). As with the Pisa results, the scoring is then transformed so that the mean is 500 and the standard deviation 100.

The database contains information for, approximately, 615,000 students: 230,000 in Grade 3, 240,000 in Grade 6 and 145,000 in Grade 10. If we take into account students whose parents responded to the questionnaire, we obtain 321,544 students: 145,096 students in Grade 3, 123,811 students in Grade 6 and 52,637 students in Grade 10. An important question concerns the impact of parents’ level of education on their child’s academic performance. We create a variables with three ‘homogenous’ categories, based on the highest level of education observed among parents:¹¹ before Grade 11, Grade 12/Vocational Training/Short-Cycle Tertiary Education and Bachelor/Master/Doctorate. Respectively, they are coded from 0 to 2 and correspond, according to the *International Standard Classification of Education* (ISCED 2011), respectively to Levels 0 to 2, Levels 3 to 5 and Levels 6 to 8. The rest of the data is described in Tables 3 to 7, in Appendix.

We illustrate in Figure 1 the impact of the parents’ highest level of education on child’s mathematics scores, for Grades 3 and 10. We plot the cumulative distribution functions of the scores, conditional on the parents (highest level of schooling) group. For each grade, we can see that

Figure 1: CDFs of mathematics scores according to parents’ highest level of education



the cumulative distribution functions are ordered in the sense of *first order stochastic dominance*:

¹¹ We could also have chosen the father’s highest level, or that of the mother. In this paper, we do not investigate whether the effects of parental schooling can be explained by assortative mating, or if the partial effects of parents can be differentiated. We consider *parents’ education as a potential* from which children can benefit. See [Holmlund et al. \(2011\)](#), Page 3, for a detailed discussion on this issue.

For any given score, the probability of having a score higher than it, is all the greater the higher the parent group. This pattern holds true for mathematics, but it also applies to all subjects and all grades in our data (discussed hereafter). First-order dominance is generally considered to be a clear indication of *inequality of opportunity*, since academic performance depends on a dimension beyond the child’s control (Lefranc *et al.*, 2009; Jaoul-Grammare and Magdalou, 2013). This result is in line with a robust trend already observed in the literature.

The second question looks at the evolution of parental influence on child’s results, throughout schooling. In Figure 1, we see a convergence of the three conditional CDFs between Grades 3 and 10 (in mathematics), suggesting that the dependence of results on parent group decreases. However, such an observation must be treated with caution. First, the *tests are of a different nature* for each grade. Then, the fact that the results are standardized (average of 500) can be misleading for comparison purposes. Finally, external factors can significantly impact the results, that can only be analyzed by econometric estimates.

One possible bias in the comparison of the impact of parents’ level of education on child’s score, at different grades, may be linked to a *composition effect*. For instance, the convergence of the CDFs between Grades 3 and 10 may be a consequence of the fact that, in the group of children whose parents have the lowest level of education, only the most gifted children remain represented in Grade 10. As established in Table 7 in Appendix, we do not observe, for each year and each grade, any significant difference in the proportions that each group represents: The proportions are all close to those obtained at global level, i.e. 9.5%, 31.5%, and 59% for parents with ISCED levels, respectively, 0–2, 3–5, and 6–8. We can therefore consider that our data do not suffer from this compositional bias.

The equivalent of Figure 1 in literature and English is provided, respectively, in Figures 9 and 10 in Appendix. Whereas the convergence of CDFs is evident in mathematics and literature between Grades 3 and 10, this is not the case for English: The inequality of opportunity observed at Grade 3 seems to increase at Grade 10. This finding is confirmed by the econometric analysis presented in the following sections.

3. Initial estimation strategy and results

3.1. Impact of parent’s highest level of education

In this section we regress students’ score in the three subjects under consideration (three regressions), focusing on the impact of parents’ highest level of education. The score of student i is denoted y_{its} , where t is the year (from 2016 to 2019) and s the school attended. As described earlier, parents’ highest level of education is modeled by an ordered categorical variable with 3

possible values, hereafter referred as *parent groups*. We introduce independent dummy variables (p_j with $j = 0, 1, 2$) that allow us to compare the impact of parent groups two by two, with the lowest group ($j = 0$) as the reference value.

We also introduce two groups of control variables, one for the *student characteristics* (each denoted s_k) and another for the *households characteristics* (each denoted h_k).¹² For each subject, we make an overall estimate of the student's score for all academic years and all grades. Although the tests are common to all schools in the Community of Madrid over the whole period, results may vary in time (from one year to another), and in space (from one school to another, particularly between public and private schools). With the aim of controlling these dimensions, we introduce *two fixed effects*, one for the academic year (a_t) and one for the identifier of the school s where the student i is registered (b_s).¹³ One obtains the following regression:

$$y_{its} = \alpha_0 + \sum_{j=1,2} \alpha_j p_{ji} + \sum_k \beta_k s_{ki} + \sum_k \gamma_k h_{ki} + a_t + b_s + \epsilon_{its}. \quad (1)$$

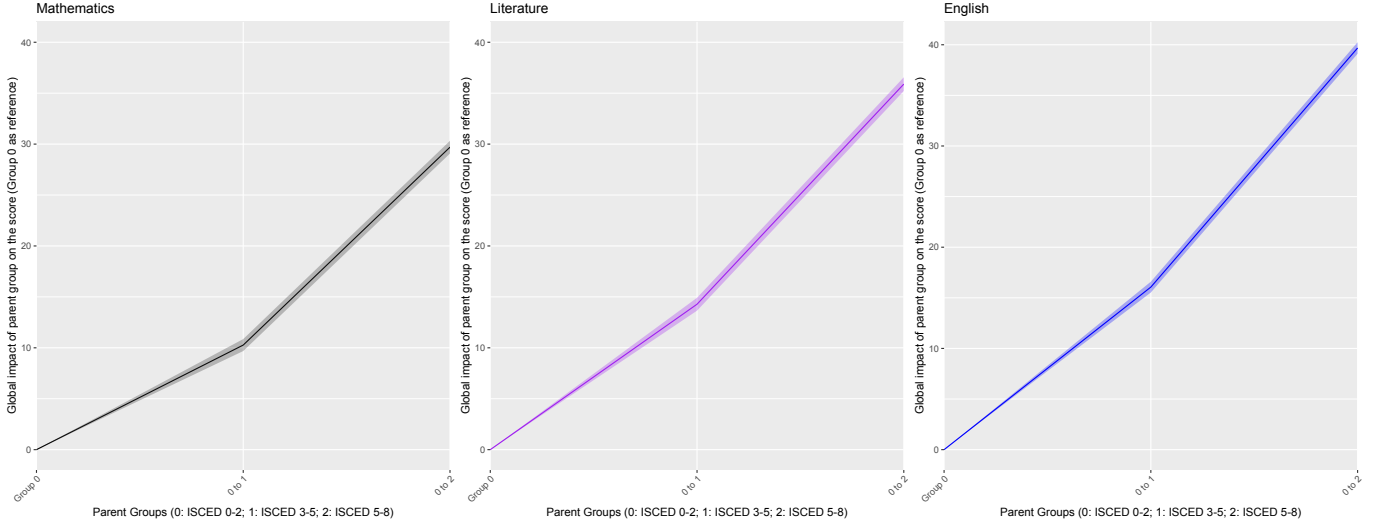
This first regression, for each subject, isolates the influence of parents' highest level of education on student's score, without distinguishing the grade. The results are shown in Table 8 for mathematics, Table 9 for literature and Table 10 for English. Whatever the regression specification (including or not the control variables), and for each subject, the results are the same and robust: moving from one parent group to a higher one, significantly increases student's score. For instance, in mathematics, moving from Parent Group 0 (ISCED 0 to 2) to Parent Group 2 (ISCED 6 to 8) increases the score by 29.7 points on average when all control variables are included (regression 3 in Table 8). The effect is larger in literature than in mathematics, and in English than in literature. We summarise these results in Figure 2 (estimates with all control variables, also plotting standard deviation). This result confirms a trend already widely observed in the literature, namely that the parents' level of education has a major influence on the child's academic performance, whatever the subject.

A more complex question (at the heart of this paper) concerns the evolution of this impact over time, throughout children's education. To this end, we supplement the previous regression with interaction components between parental highest education and grade level of the student (Grade 3, Grade 6 or Grade 10). Precisely, we compare the grades two by two (3 vs. 6, 6 vs. 10

¹² Student characteristics include country of birth, number of days a week spent on homework, and gender. Household characteristics include country of birth of both parent, the frequency of use of books/computers/internet at home, the number of books at home, the employment status of both parents, the frequency with which parents talk about school subjects/teach homework/help with homework/check homework with the child.

¹³ At this stage, we do not introduce a fixed effect for the grade because the mean score, whatever the grade, is always normalised at 500 at the population level (and the standard deviation at 100). Even if there is a small variation in average scores between subjects and grades in our samples, they remain very close to 500. Notice that the constants estimated in regressions (1) is close to this value (see results tables).

Figure 2: Global impact of parents' highest level of education on student's score



Reading: In mathematics, moving from Parent Group 0 (Education ISCED 0–2) to Parent Group 1 (Education ISCED 3–5) increases the student's score by about 10 points (which represents 2% of the normalised average score and 10% of the normalised standard deviation, respectively equal to 500 and 100). $\pm$ standard deviation around the curve.

and 3 vs. 10), running a regression for each possible grades pair (u, v) , where $u < v$. In each of these regressions, we retain only the observations of students' scores for grades u and v (excluding the third grade). With g indicating the grade, we introduce a dummy variable $I(u, v)$ which takes the value 0 if $g = u$ (reference grade), and 1 if $g = v$. When we compare the grades u and v , we obtain the following expression, with $g \in \{u, v\}$:

$$y_{igts} = \alpha_0 + \sum_{j=1,2} \alpha_j p_{ji} + \eta I(u, v) + \sum_{j=1,2} \theta_j p_{ji} I(u, v) + \sum_k \beta_k s_{ki} + \sum_k \gamma_k h_{ki} + a_t + b_s + \epsilon_{igts}. \quad (2)$$

The η coefficient indicates the extra points obtained on average by the students in Grade v , compared with Grade u , whatever the parent group. As the average score is standardised at 500 for all grades, this coefficient is not very informative (values other than 0 are due to sample selection). We focus on coefficients α_j and θ_j , for $j = 1, 2$. By definition of p_{ji} and $I(u, v)$, the variable $p_{ji} \times I(u, v)$ takes the value 1 if and only if student i is in Grade v , with parents from group j . Since the parent reference group is the lowest ($j = 0$), and the same applies to the grade (the reference is u , with $u < v$), the α_j coefficient is interpreted as the extra points obtained by the student in Grade u when the parents are in group j (compared with 0), and θ_j is interpreted as the marginal impact of parent's group j , when the student is in Grade v instead of Grade u .

For each subject, we obtain 9 regressions (3 regressions, depending on whether the student and household characteristics are included or not, for each of the 3 pairs of grades compared). The results are shown in Tables 11, 12 and 13 for, respectively, the comparison of Grade 3 versus

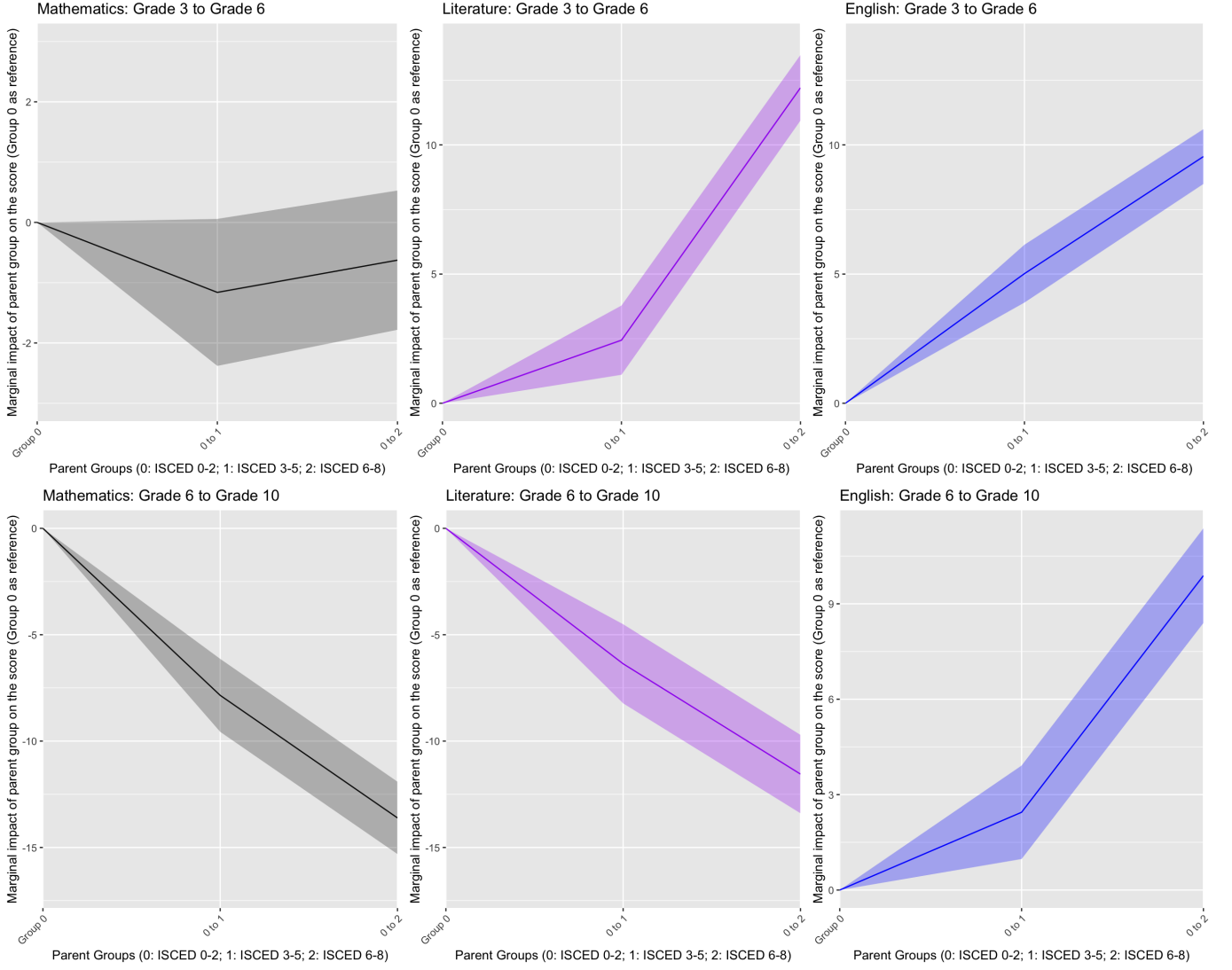
Grade 6, Grade 6 versus Grade 10, and Grade 3 versus Grade 10 (each table showing the results for the 3 subjects). As established in the regressions without interaction components, parents' highest level of education significantly increases student's score, in each subject (coefficients α_j). If we focus on the regressions that take into account all the control variables (columns 3, 6 and 9 for each subject in, respectively, Tables 11, 12 and 13), we can see that moving from Parent Group 0 (ISCED 0 to 2) to Group 1 (ISCED 3 to 5) significantly increases the score, between 11 and 16 points in the three subjects. From Parent Group 0 to 2 (ISCED 6 to 8), the score increases between 28 and 37 points. These last values are relatively high, in the range of a third of the normalised standard deviation at the population level (equal to 100).

We now examine the estimates of the interaction components (θ_j), again focussing on the regressions that take into account all the control variables (columns 3, 6 and 9 for each subject, of Tables 11, 12 and 13). In mathematics, there is no significant marginal impact of parent group between Grades 3 and 6, but the impact is significant and negative between Grades 6 and 10 (and, consequently, also between Grades 3 and 10). As a result, the impact of parents' highest level of education diminishes as student progresses through grades, from Grade 6 upwards. Between Grades 6 and 10, moving from Parent Group 0 to 1 results in a 7.9 point decrease in the score gap, and moving from Parent Group 0 to 2 results in a 13.6 point decrease in the score gap. A comparable pattern applies to literature, with a small difference: Between Grades 3 and 6, the marginal impact of moving from Parent Group 0 to 2 is significantly positive (about 12.2 points). The same change in parent group, but between Grades 6 and 10, implies a gap reduction of 11.6 points. The effect is therefore a bell-shaped curve in literature. The situation is completely different in English. Between Grades 3 and 6, and also between Grades 6 and 10, the impact of parent group increases significantly. For instance, between Grades 3 and 6, the marginal impact of moving from Parent Group 0 to 2 is around 9.5 points and, from Grade 6 to Grade 10, is about 9.9 additional points. Hence, influence of parent group is continuously and significantly increasing in English. We summarise the main results of the estimates with interaction components in Figure 3 (estimates with all control variables).

3.2. Impact of parental investment and student's effort

In this paper, we analyse the impact of parental environment on academic performance for their child through two channels: parent's level of education (thus, skills acquired before bringing up children), and parental investment (an effort made by parents during the upbringing of their child). In the previous section, we showed that the effect of parents' education tends to decrease between Grades 6 and 10 in mathematics and literature, but increases in English. This section focuses on the second aspect, i.e. the impact of parental investment, but also on the impact of efforts made

Figure 3: Marginal impact of parents' highest level of education on student's score

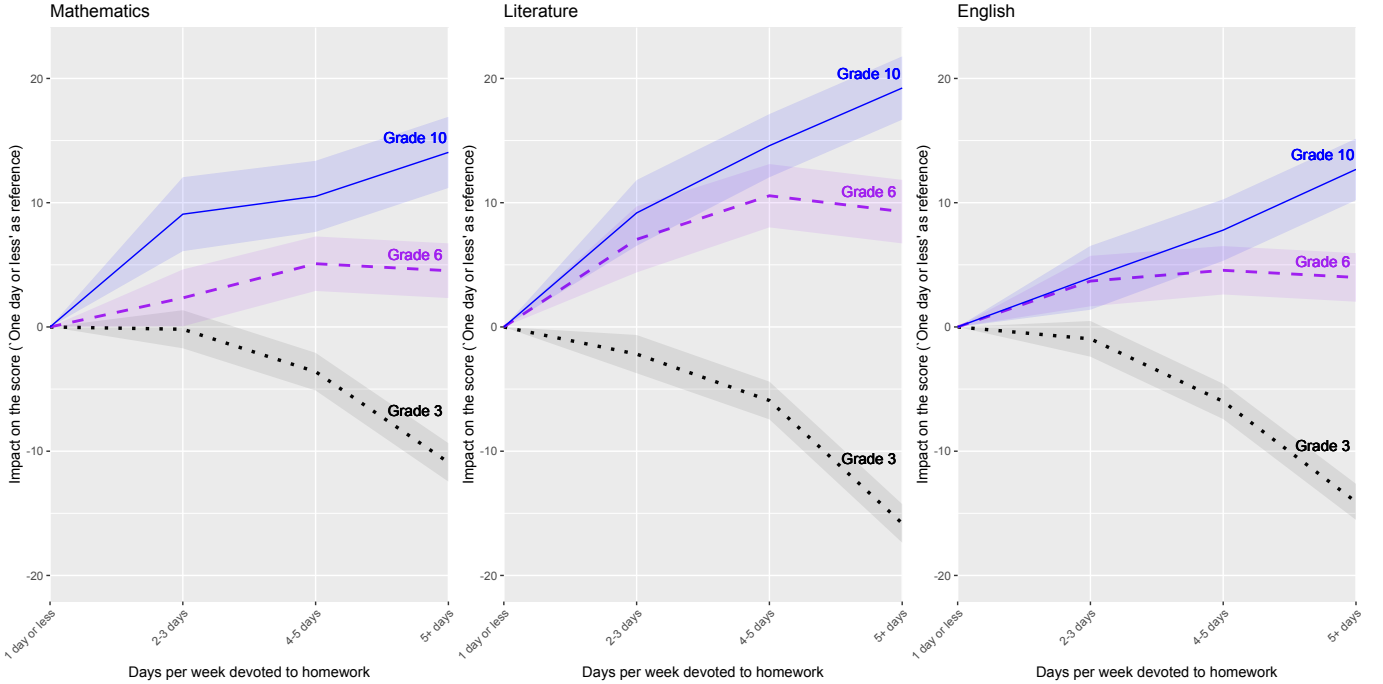


Reading: Between Grade 6 and Grade 10 in mathematics, the marginal impact on the student's score, of Parent Group 1 (ISCED 3 to 5) compared with Parent Group 0 (Education ISCED 0–2), is reduced by about 7.5 points (which represents 1.5% of the normalised average score and 7.5% of the normalised standard deviation, respectively equal to 500 and 100). $\pm$ standard deviation around the curve.

by the child in her studies. We run regressions comparable to Equation (1), but one per subject and per grade. The results are presented in Table 14. We focus on two explanatory dimensions for, respectively, parental investment and child’s effort: ‘frequency parents talk to their child about school’ and ‘days per week devoted to homework’ (both are introduced in the form of dummy variables).

Estimates are summarised in Figure 4. In terms of student’s effort, a clear trend emerges, whatever the subject: in Grade 3, the more days per week devoted to homework, the lower the scores (compared with the lowest category, i.e. ‘one day or less’). In Grade 6, an increase in the number of days dedicated to homework goes with an increase in results up to 4-5 days per week, then decreases thereafter. In Grade 10, scores increase with the number of days dedicated to homework. The negative impact of the number of days dedicated to homework in the first grade

Figure 4: Impact of student’s effort on score



Reading: In Grade 3 in mathematics, spending more than 5 days a week on homework, as opposed to one day or less, reduces a student’s score by around 10 points (which represents 2% of the normalised average score and 10% of the normalised standard deviation, respectively equal to 500 and 100). $\pm$ standard deviation around the curve.

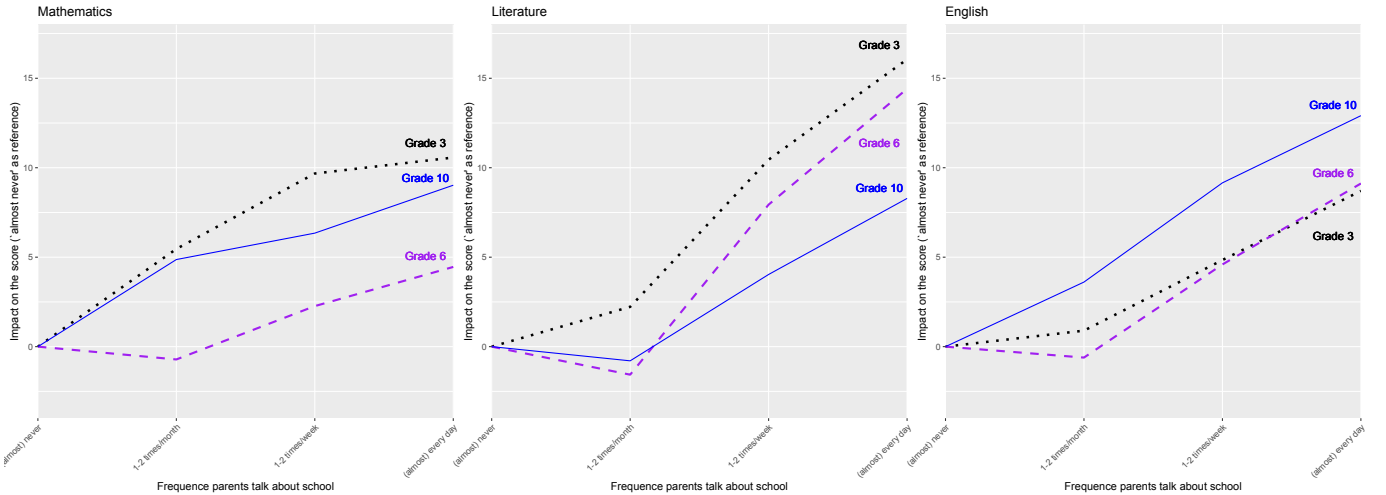
seems to indicate a reverse causality (which could be confirmed by further investigations), i.e. it is poor school results that implies more time devoted to homework (the volume of homework being lower than in the higher grades). As students progress in their studies, effort seems to have an increasing impact, culminating in a clear positive impact in Grade 10. Finally, if we focus on the highest effort category (‘five days or more’) in Grade 10, we can see that the greatest positive impact is for literature (+19.23 points), then mathematics (+14.05 points), and finally English

(+12.68 points).

The number of days per week dedicated to homework cannot be considered as a purely effort variable for the child, as homework is supervised by parents (particularly in early childhood). However, children gain independence as they progress through the grades and, as time goes on, the more work they do, the better their results. This form of emancipation seems to be weaker in English, compared with the other two subjects.

The impact of parental investment on child’s achievement also seems to go hand in hand with lower emancipation in English. First of all, we observe in Table 14 that, whatever the subject or grade (with some few exceptions), if the frequency with which parents talk to their child about school increases, then student achievement improves. However, between Grades 3 and 10, if we focus on the highest category of investment (‘every or almost every day’), the positive impact of parental investment decreases slightly in mathematics (from 10.6 to 9.0 points) and sharply in literature (from 16.1 to 8.3 points), while it rises sharply in English (from 8.7 to 12.9 points). Hence, between Grades 3 and 10, the impact of parental investment is clearly increasing in English, unlike in the the two other subjects. These results are shown in Figure 5 (where standard deviations have not been shown for the sake of clarity).

Figure 5: Impact of parental investment on student’s score



Reading: In Grade 10 in mathematics, if the parents talk to their child about school ‘once or twice a month’ instead of ‘never’, the child’s score increases by around 5 points (which represents 1% of the normalised average score and 5% of the normalised standard deviation, respectively equal to 500 and 100).

Finally, if we look at all the explanatory variables in Grade 10 (Table 14), we see that in mathematics and literature, the impact of child’s effort exceeds the impact of parental investment and is comparable to the impact of parents’ level of education. This is not the case in English: the impact of the child’s effort and that of the parents’ investment are roughly comparable, while

the impact of parents' level of education is much greater.

4. Addressing endogeneity: Instrumental variable (IV) analysis

4.1. Identification strategy

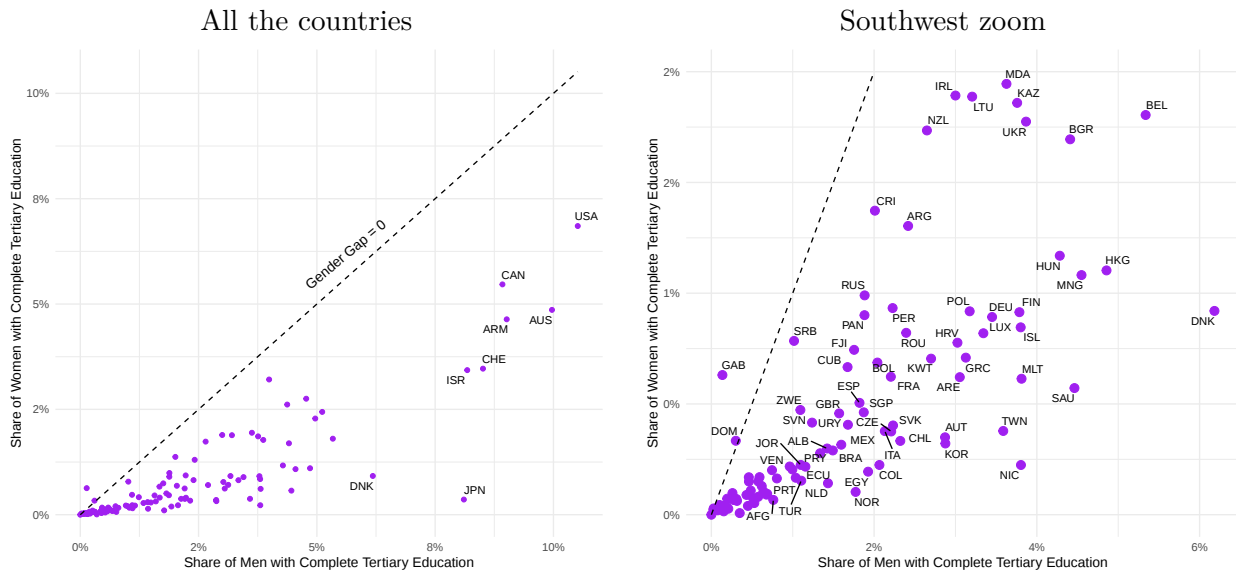
Section 3 relies on linear regressions to estimate the relationship between parental environment and student achievement. The advantage is that they allow for detailed analysis, as in Figure 3 where, from Grade to Grade and by subject, we can see the differentiated impact of the three categories of parental education. However, these estimates may be biased due to the endogeneity of the dependent variables of interest. For instance, parents' level of education is likely correlated with unobserved family characteristics, such as motivation or inherited abilities, which also support a child's accumulation of human capital. This correlation can make it difficult to isolate the true causal effect of parental schooling.

To address this endogeneity issue and identify the causal impact of parents' level of education on the academic performance of students, we employ an instrumental variable approach. Following [Cordero *et al.* \(2018\)](#), who recommend the use of historical sources of exogenous variation, our chosen instrument is the gender gap in tertiary education in 1960 in the country of birth of the more educated parent. This variable is constructed using the Educational Attainment Data from [Barro and Lee \(2013\)](#). The gender gap is specifically calculated as the difference between the share of women and men who completed tertiary education. The viability of this instrument rests on its significant variation across the countries relevant to our analysis, a fact illustrated in Figures 6 and 7. The parents in our database come from 115 different countries, out of the 146 studied by [Barro and Lee \(2013\)](#) in 1960. For these countries, Figure 6 plots the share of women versus the share of men with complete tertiary education, demonstrating substantial cross-country differences. Figure 7 further details this variation by ranking countries according to the gender gap ('share of women' minus 'share of men'), showing that the gap is negative for almost all the countries we are interested in, indicating lower rates of higher education for women.

Our identification strategy relies on the assumption that the gender gap in the parent's country of origin influences their child's academic performance exclusively through its effect on that parent's educational attainment. While this assumption cannot be tested directly, the richness of our data allows us to mitigate most concerns. For instance, one potential issue is that a country's wealth might be correlated with its gender gap, and that the wealth of the parents' country of origin could independently influence a child's academic success. The association between the gender gap in tertiary education and wealth is not confirmed in the literature. By analyzing global educational gender gaps (not only in tertiary education), [Minasyan *et al.* \(2019\)](#) show a

PARENTAL ENVIRONMENT AND STUDENT ACHIEVEMENT

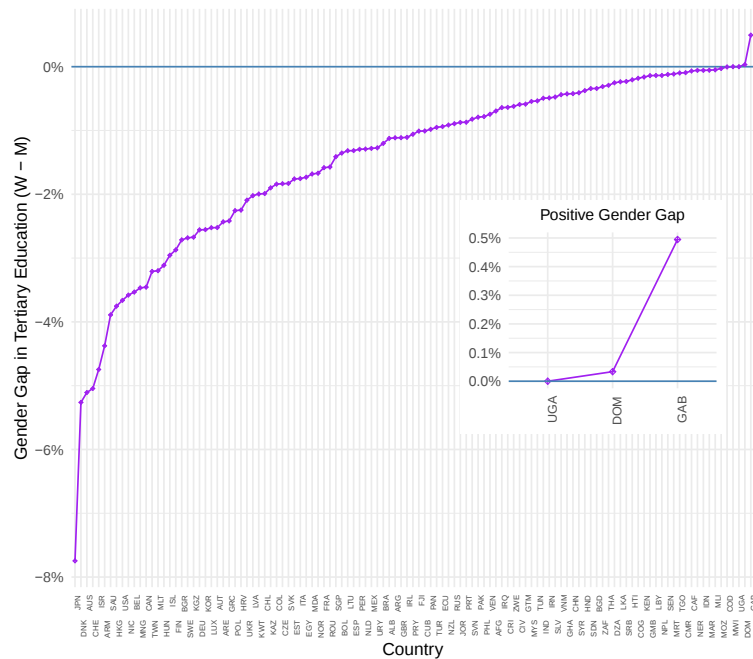
Figure 6: Disparities in tertiary education between women and men in 1960



Reading: In 1960, in Australia (AUS), approximately 5% of adult women had completed higher education, compared to approximately 10% of adult men.

Source: Educational Attainment Data, [Barro and Lee \(2013\)](#).

Figure 7: Gender gaps in tertiary education in 1960



Reading: In 1960, in Gabon (GAB), the share of women with complete tertiary education was 0.5% higher than that of men.

Source: Educational Attainment Data, [Barro and Lee \(2013\)](#).

relative weak correlation of 0.25 with economic growth, and that there is no significant difference when considering per capita income measures instead of growth. Focusing on the gender gap in tertiary education, [Huber and Paule-Paludkiewicz \(2024\)](#) show that it can be negative or positive regardless of wealth, and establish that the difference between countries is partially explained by differences in social norms, relating to gender, within each country.¹⁴

The causal effect of the parent’s highest level education on child’s score is estimated using a two-stage least squares (2SLS) model. We first regress the endogenous variable, the parents’ highest level of education (PHE, which can take three values, namely 0, 1 and 2 for, respectively, ISCED 0–2, 3–5, and 6–8), on the instrument, the gender gap in tertiary education (GG), including all the exogenous controls presented in regression (1). As in Figure 7, the gender gap is calculated as the difference between the share of women and the share of men who have completed higher education. The first-stage equation is:

$$\text{PHE}_i = \pi_0 + \pi_1 \text{GG}_i + \sum_k \sigma_k s_{ki} + \sum_k \lambda_k h_{ki} + a_t + b_s + \nu_{its} . \quad (3)$$

A negative value for π_1 can be expected, which would mean that the greater the gender inequality in education in the country of origin, the more educated parents tend to be. The literature explains this pattern through selective migration. For foreign parents, most of whom come from developing countries in our study, migration can be perceived as an opportunity.¹⁵ Because of mobility costs and spatial frictions, it is typically the most educated who are able to migrate ([Schmutz et al., 2021](#); [Bergman et al., 2024](#)). This pattern is, for instance, confirmed in [Black et al. \(2006\)](#) on the basis of three concrete case studies (Kerala in India, Bangladesh, and Albania).

We then use the predicted values of parents’ highest level of education ($\widehat{\text{PHE}}_i$) from the first stage to estimate the impact on student scores (y_{its}). The second-stage equation is:

$$y_{its} = \alpha_0 + \alpha_1 \widehat{\text{PHE}}_i + \sum_k \beta_k s_{ki} + \sum_k \gamma_k h_{ki} + a_t + b_s + \epsilon_{its} . \quad (4)$$

In this framework, the coefficient α_1 represents the causal effect of parents’ highest level of education on student academic achievement. This regression is therefore equivalent to Equation (1), but with an instrumented version of parents’ highest level of education, and without it being put into dummy variables (which would have required a larger number of instruments, which could

¹⁴ For example, the United States, which in 1960 was the richest country in the world, had the 9th most unequal gender gap in completed tertiary education in our database, among the 115 countries observed.

¹⁵ By examining the Great Migration (1940–1970), during which millions of African Americans left the segregationist South for cities in the North, [Derenoncourt \(2022\)](#) challenges the notion that moving to areas offering better opportunities systematically guarantees greater intergenerational social mobility.

have posed potential problems in terms of validity).

As in Section 3, we complete the previous regression with interaction components between parental highest education and the student’s grade level. Again, we compare the grades two by two (3 vs. 6, 6 vs. 10 and 3 vs. 10), running a regression for each possible grades pair (u, v) where $u < v$, and excluding the non-concerned third grade. With g indicating the grade, we introduce a dummy variable $I(u, v)$ which takes the value 0 if $g = u$ (reference grade), and 1 if $g = v$. When we compare the grades u and v , we obtain the following expression, with $g \in \{u, v\}$:

$$y_{igts} = \alpha_0 + \alpha_1 \widehat{\text{PHE}}_i + \eta I(u, v) + \alpha_2 \widehat{\text{PHE}}_i I(u, v) + \sum_k \beta_k s_{ki} + \sum_k \gamma_k h_{ki} + a_t + b_s + \epsilon_{igts}. \quad (5)$$

This regression is the instrumented version of Equation (2). Again the η coefficient—which indicates the extra points obtained on average by the students in Grade v , compared with Grade u , whatever the parent group—is not informative due to the normalisation of the scores. The coefficient we are interest in is α_2 , which corresponds to the average impact, from Grade u to Grade v , of a one-level increase in the highest level of parental education (among the three possible levels).

4.2. Results

The results of the instrumental variable estimations of Equations (3) and (4) for the three subjects (mathematics, literature and English), aggregating all grade levels, are presented in Table 1. The first-stage results show a strong and significant relationship between our instrument, the gender gap in tertiary education (GG), and the endogenous variable, parents’ highest level of education (PHE). The coefficient on the gender gap is negative and statistically significant across all three subjects (approximately -0.05). This suggests that a larger gender gap in the parents’ country of origin, which indicates lower educational attainment for women, is associated with a higher level of education for the parents in our sample, consistent with a selective migration pattern. The F-statistics for the excluded instrument are well above conventional thresholds for weak instruments, with values of 248.06 for mathematics, 242.11 for literature, and 250.91 for English, confirming the instrument’s strength. The reduced-form estimates confirm a statistically significant negative relationship between the gender gap and children’s scores across all subjects.

The second-stage results in Table 1 reveal a positive and significant causal effect of parents’ highest education on their children’s academic scores in all subjects. An increase in the parents’ education level (recalling that we have three possible values, 0, 1 and 2) leads to a score increase of 37.46 points in mathematics, 51.30 points in literature, and 72.47 points in English. These results confirm that the parental environment has a substantial causal influence on student achievement after addressing potential endogeneity (persistent effect). These results also establish that the

Table 1: Global and causal impact (IV) of parents' highest level of education on student's score

	(1) Mathematics	(2) Literature	(3) English
First stage: Impact of gender gap in tertiary education (GG) on parents' highest level of education (PHE)			
Gender gap in tertiary education	-0.051*** (0.003)	-0.051*** (0.003)	-0.052*** (0.003)
F-test of excluded instrument	248.06***	242.11***	250.91***
Reduced form: Impact of GG on student's score			
Gender gap in tertiary education	-1.86*** (0.533)	-2.634*** (0.568)	-3.764*** (0.487)
Second stage: Impact of PHE, instrumented by GG, on student's score			
Parents highest education	37.46*** (10.46)	51.30*** (11.36)	72.47*** (9.89)
Academic Year FE	Yes	Yes	Yes
School FE	Yes	Yes	Yes
Child's characteristics	Yes	Yes	Yes
Household's characteristics	Yes	Yes	Yes
Observations	297491	299350	297851
Standard errors in parentheses			
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$			

impact appears to be stronger in English, which confirms our observations in Section 3.

Table 2 presents, still on the basis of Equations (3) and (4), the results disaggregated by grade level, allowing us to observe the evolution of this causal impact. In all grade-level regressions, the first-stage F-statistics are robust, ranging from 65.45 to 99.80, which supports the validity of the instrument across all subsamples. Now we focus on analyzing the results of the second-stage. One clear fact is the similar trend in mathematics and literature. The causal effect of parents' highest education is positive and significant in Grade 3 (at 1%) and Grade 6 (at 5%), but decreasing between the two grade levels. By Grade 10, the effect becomes statistically insignificant. In contrast to the other subjects, the causal effect of parents' education on English scores remains positive and statistically significant across all three grades (always at 1%). While the magnitude of the coefficient decreases between Grade 6 and Grade 10, the influence of parental education persists strongly into adolescence for foreign language learning. As an illustration, in Grade 10, an increase in the parents' education level leads, on average, to an insignificant increase of about 10 points in mathematics and literature, and a significant increase of more than 50 points in English. This pattern seems to confirm that the causal influence of parental education on mathematics and literature performance diminishes as the child gets older, while it increases in English (Matthew effect), as established in Section 3.

Table 2: Marginal and causal impact (IV) of parents' highest level of education on student's score, by grade level

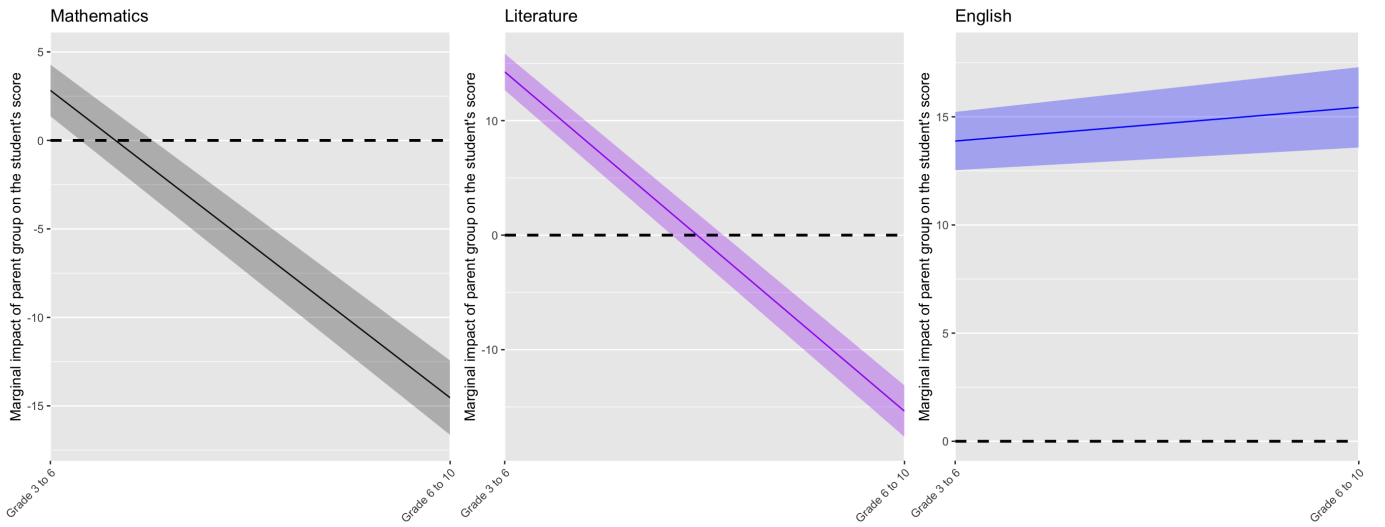
	Mathematics			Literature			English		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Gr.3	Gr.6	Gr.10	Gr.3	Gr.6	Gr.10	Gr.3	Gr.6	Gr.10
First stage: Impact of gender gap in tertiary education (GG) on parents' highest level of education (PHE)									
Gender gap in tertiary education	-0.040***	-0.057***	-0.065***	-0.040***	-0.056***	-0.063***	-0.041***	-0.057***	-0.066***
	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)
F-test of excluded instrument	65.45***	99.60***	90.25***	66.91***	97.22***	84.09***	67.24***	99.80***	91.58***
Reduced form: Impact of GG in Education on student's score									
Gender gap in tertiary education	-2.763***	-1.924**	-0.859	-3.706***	-2.492**	-0.601	-3.229***	-4.619***	-3.609***
	(0.790)	(0.896)	(1.109)	(0.795)	(1.029)	(0.981)	(0.743)	(0.802)	(0.968)
Second stage: Impact of PHE, instrumented by GG in Ed., on student's score									
Parents highest education	72.65***	33.65**	13.40	93.22***	41.84**	11.03	82.58***	80.15***	54.05***
	(21.04)	(15.94)	(17.13)	(21.83)	(18.57)	(15.65)	(19.87)	(15.27)	(15.00)
Academic Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	133592	113824	51761	133899	113876	51575	132695	113472	51684

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We now discuss the estimation results of Equation (5), which includes an interaction term between parental highest education and the student's grade level. These results are summarised in Tables 15, 16 and 17 for, respectively, the transition from Grade 3 to Grade 6, Grade 6 to Grade 10, and Grade 3 to Grade 10. Once again, the causal effect of the variable 'parents' highest level of education' remains positive and significant, regardless of the subject and regardless of the econometric specification. We are now interested in the marginal impact from one grade to another, considering only estimates that include all control variables. These results can be viewed in Figure 8.

Figure 8: Marginal and causal impact (IV) of parents' highest level of education on student's score



Reading: Between Grades 3 and 6 in literature, increasing the parents' highest level of education (among the three possible categories: ISCED 0–2, ISCED 3–5 and ISCED 6–8) by one category increases the score by an average of 14.26 points. Thereafter, between Grade 6 and Grade 10 in literature, the marginal impact of a higher category for the parents' highest level of education, on the student's score, is reduced by an average of 15.37 points. $\pm$ standard deviation around the line.

In mathematics, between Grades 3 and 6, a one-category increase in the parent's level of education leads, on average, to a slight increase (significant at only 10%) of 2.8 points of the student's score. Then, between Grades 6 and 10, this impact is significantly negative (-14.5 points). The trend is similar in literature, except that the positive impact is significant between Grades 3 and 6 (+14.3 points). As in mathematics, it is then negative and significant between Grades 6 and 10 (-15.4 points). The results for these two subjects confirm a general downward trend in the impact of parents' level of education on student's score in mathematics between the ages of 8 and 15, and a bell-shaped trend in literature. The pattern in English is completely different. The marginal impact remains strongly positive (and significant) between Grades 3 and 6 (+13.9 points), and Grades 6 and 10 (+15.5 points). Overall, between Grades 3 and 10, the

marginal impact decreases in mathematics and literature (by -4.7 and -2.6 points, respectively) while it increases by 32.5 points in English.

To sum up, these instrumental variable estimates reinforce the findings from our initial OLS models, providing stronger causal evidence for a mixed Matthew effect: the influence of the parental environment diminishes for mathematics and literature as students mature (with a bell-shaped curve in literature), but it is unambiguously amplified in foreign language acquisition.

5. Discussion, related literature, and policy implications

Matthew effect in education. Few empirical studies have attempted to test the Matthew effect hypothesis in education, most of them applied to reading abilities and its impact on acquisition of literacy (and other related skills). They all agree that the differences in reading abilities in the early education stages continue on until adulthood (Cunningham and Stanovich, 1997; Rigney, 2010), but the results are mixed on the existence of a Matthew effect: There is not a strong support for a pattern of widening or decreasing achievement differences (Pfof *et al.*, 2014). Some papers find that the effect is strongly increasing (Awaida and Beech, 1995; Howley, 2001), others that it is intermediate (Bast and Reitsma, 1998) and some that it is not even significant or related to social background (Shawitz *et al.*, 1995; Protopapas *et al.*, 2011). But no study has tested, on a unified database, the existence of a possible Matthew effect on the core subjects of mathematics, literature and the main foreign language (in our case English).

While our study confirms a persistent impact of parental environment on child's academic performance (a fact widely accepted since the Coleman report, 1966), the results differ between subjects as regards the Matthew effect. From age 11 to 15, the effect of parent's level of education decreases in mathematics and literature, while it increases in English. At age of 15, spending '5 days or more' doing homework (compared with 'one day or less') increases child achievement, but more strongly in literature than in mathematics, and more strongly in mathematics than in English. Similarly, at the same age, the impact of 'the frequency with which parents talk to their child about school' is stronger in English, followed by mathematics and then by literature (although the order is reversed at age 8).

To sum up, these results therefore reflect a partial emancipation (from the influence of parental environment) in mathematics and literature, while social determinism increases in English. In the first two subjects, the results echo the work of cognitive psychology initiated by Jean Piaget, according to which the child is partly master of his or her own development: They have an intrinsic ability to learn, without this necessarily being transmitted by others, and their strategies and involvement play a role in their academic performance (Onatsu-Arvilomni and Nurmi, 2000).

The notion of an age of consent (for responsible choice) can therefore make sense, including in education, and this age can be set between 12 and 16, as proposed by [Hufe et al. \(2017\)](#).

Conversely, our results confirm the social dimension of learning a foreign language, compared with other academic subjects. In that case, external factors appear to play a decisive role in the learning process ([Vygotsky, 1978](#)). In addition, the theory of [Cunha and Heckman \(2007\)](#) which describes a path-dependency in the formation of cognitive skills, seems to be confirmed. As [Gardner \(1968\)](#) and [Dornyei \(1998\)](#) point out, the acquisition of a new language involves a great deal of integrative motivation (in the sense that people are interested in learning a language because they want to communicate with the other language community), and parents play a crucial role in encouraging this integrative motivation (as opposed to instrumental motivation). According to [Gardner \(1968\)](#), parents play two roles in their child's success in learning a second language: an active role which consists of actively and consciously encouraging their child to learn the language, and a (more important) passive role, which consists of the attitudes that parents have towards the community whose language their child is learning.

Policy implications. Our results, which need to be confirmed using other databases and complementary methodologies, have several implications for educational policies. The first concerns remediation programmes, aimed at improving the skills of children experiencing difficulties. The second concerns national selection processes in higher education ('Grandes Ecoles' in France, for instance), which include foreign language skills as a criterion for admission.

With regard to remedial programmes, the main recommendation resulting from the empirical estimates of [Cunha and Heckman \(2007\)](#) and [Cunha et al. \(2010\)](#) is to focus (adolescent) remediation strategies for disadvantaged children on the development of non-cognitive skills. Our results indicate that programmes focusing on cognitive skills can also be effective, particularly in mathematics. These results are in line with those of [Bahr \(2007, 2008\)](#), who assesses US postsecondary remediation programmes. He first observes that the degree of deficiency (depth) and the number of deficient basic skill areas (breadth) are good predictors of successful math remediation: Those who require the least remediation are the most likely to remediate successfully. But he also observes that when remediation works (at a low rate, unfortunately) it works extremely well: 'students who remediate successfully in mathematics exhibit attainment that is comparable to that of students who achieve college mathematics skill without the need for remediation' ([Bahr, 2008](#), Page 442).

The second implication in terms of public education policy concerns the weight of foreign languages in the selection process to access higher education, at different levels. We have found that parents' education and involvement are essential factors in children's success in learning a foreign language. What's more, the child's effort has (slightly) less impact on results than in

subjects such as mathematics or literature. This is a strong sign of inequality of opportunity, and including foreign languages as an admission criterion reinforces this inequality. It therefore seems essential that educational systems give a high priority to foreign language teaching, particularly in non-English-speaking countries, from an early age, so as not to further penalise children from socially disadvantaged backgrounds.

References

- AWAIDA, M. and BEECH, J. (1995). Children's lexical and sublexical development while learning to read. *Journal of Experimental Education*, **63**, 97–113.
- BAHR, P. (2007). Double jeopardy: Testing the effects of multiple basic skill deficiencies on successful remediation. *Research in Higher Education*, **48**(6), 695–725.
- (2008). Does mathematics remediation work?: A comparative analysis of academic attainment among community college students. *Research in Higher Education*, **49**(5), 420–450.
- BARRO, R. and LEE, J.-W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, **104**, 184–198.
- BAST, J. and REITSMA, P. (1998). Analyzing the development of individual differences in terms of matthew effects in reading: Results from a dutch longitudinal study. *Developmental Psychology*, **34**(6), 1373–1399.
- BERGMAN, P., CHETTY, R., DELUCA, S., HENDREN, N., KATZ, L. F. and PALMER, C. (2024). Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. *American Economic Review*, **114** (5), 1281–1337.
- BINGLEY, P., CAPPELLARI, L. and TATSIRAMOS, K. (2022). Parental assortative mating and the intergenerational transmission of human capital. *Labor Economics*, **77**, 102047.
- BLACK, R., NATALI, C. and SKINNER, J. (2006). *Migration and inequality*. Washington: World Bank.
- CORDERO, J. M., CRISTÓBAL, V. and SANTÍN, D. (2018). Causal inference on education policies: A survey of empirical studies using PISA, TIMSS and PIRLS. *Journal of Economic Surveys*, **32** (3), 878–915.
- CUNHA, F. and HECKMAN, J. (2007). The technology of skill formation. *American Economic Review P&P*, **97**, 31–47.
- , — and SCHENNACH, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, **78**(3), 883–931.
- CUNNINGHAM, A. and STANOVICH, K. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, **33**(6), 934–945.

- DERENONCOURT, E. (2022). Can you move to opportunity? evidence from the great migration. *American Economic Review*, **112** (2), 369–408.
- DORNYEI, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, **31**(1), 117–135.
- EDWARDS, R., GIBSON, R., HARMON, C. and SCHURER, S. (2022). First in their families at university: Can non-cognitive skills compensate for social origin? *Economics of Education Review*, **91**, 102318.
- EIKA, L., MOGSTAD, M. and ZAFAR, B. (2019). Educational assortative mating and household income inequality. *Journal of Political Economy*, **127**(6), 2795–2835.
- GARDNER, R. (1968). Attitudes and motivation: Their role in second-language acquisition. *TESO Quaterly*, **2**(3), 141–150.
- HANUSHEK, E., JACOBS, B., SCHWERDT, G., VAN DER VELDEN, R., VERMEULEN, S. and S., W. (2021). The intergenerational transmission of cognitive skills: An investigation of the causal impact of families on student outcomes, ISA Discussion Paper No.14854.
- HOLMLUND, H., LINDHAL, M. and PLUG, E. (2011). The causal effect of parents' schooling on children's schooling: A comparison of estimation methods. *Journal of Economic Literature*, **49**(3), 615–651.
- HOWLEY, C. (2001). The matthew principle: A west virginia replication? *Education Policy Analysis Archives*, **3**(18).
- HUBER, S. J. and PAULE-PALUDKIEWICZ, H. (2024). Gender norms and the gender gap in higher education. *Labour Economics*, **87**, 102491.
- HUFE, P., PEICHL, A., ROEMER, J. and UNGERER, M. (2017). Inequality of income acquisition: The role of childhood circumstances. *Social Choice and Welfare*, **49**, 499–544.
- JAUL-GRAMMARE, M. and MAGDALOU, B. (2013). Opportunities in higher education: An application to france. *Annals of Economics and Statistics*, **111-112**, 295–325.
- LEFRANC, A., PISTOLESI, N. and TRANNOY, A. (2009). Equality of opportunity and luck: Definitions and testables conditions, with an application to income in France. *Journal of Public Economics*, **93**, 1189–1207.
- MASTERS, G. and WRIGHT, B. (1997). The partial credit model. In W. van der Linden and R. Hambleton (eds.), *Handbook of Modern Item Response Theory, Chapter 6*, New York: Springer Science and Business Media, LLC, pp. 101–122.
- MERTON, R. (1968). The matthew effect in science. *Science*, **159**, 56–63.
- MINASYAN, A., ZENKER, J., KLASSEN, S. and VOLLMER, S. (2019). Educational gender gaps and economic growth: A systematic review and meta-regression analysis. *World Development*, **122**, 199–217.
- ONATSU-ARVILOMNI, T. and NURMI, J. (2000). The role of task-avoidant and task-focused

- behaviors in the development of reading and mathematical skills during the first school year: A cross-lagged longitudinal study. *Journal of Educational Psychology*, **92**(3), 478–491.
- PFOST, M., HATTIE, J., DÖRFLER, T. and ARTELT, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on matthew effects in reading. *Review of Educational Research*, **84**(2), 122–136.
- PROTOPAPAS, A., SIDERIDIS, G. D., MOUZAKI, A. and SIMOS, P. G. (2011). Matthew effects in reading comprehension: Myth or reality? *Journal of Learning Disabilities*, **44**(5), 402–420.
- RIGNEY, D. (2010). *The Matthew Effect: How Advantage Begets Further Advantage*. New York: Columbia University Press.
- ROEMER, J. and TRANNOY, A. (2016). Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, **54**(4), 1288–1332.
- SACERDOTE, B. (2007). How large are the effects from changes in family environment? A study of Korean American adoptees. *Quarterly Journal of Economics*, **122**, 119–157.
- SCHMUTZ, B., SIDIBÉ, M. and VIDAL-NAQUET, E. (2021). Why are low-skilled workers less mobile? the role of mobility costs and spatial frictions. *Annals of Economics and Statistics*, **142**, 283–304.
- SHAWITZ, B., HOLFORD, T., HOLAHAN, J., FLETCHER, J., STUEBING, K., FRANCIS, D. and SHAYWITZ, S. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly*, **30**(4), 613–625.
- SUHONEN, T. and KARHUNEN, H. (2019). The intergenerational effects of parental higher education: Evidence from changes in university accessibility. *Journal of Public Economics*, **176**, 195–217.
- VYGOTSKY, L. (1978). *Mind in society: The development of higher psychological processes*. Massachusetts: Harvard University Press.

A. Appendix

Figure 9: CDFs of literature scores according to parents' highest level of education

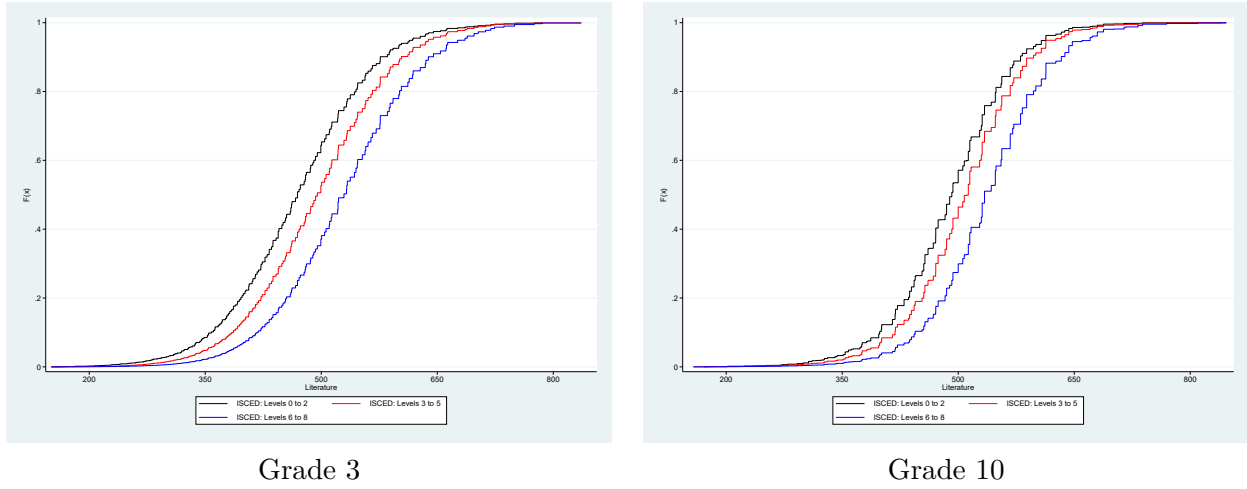
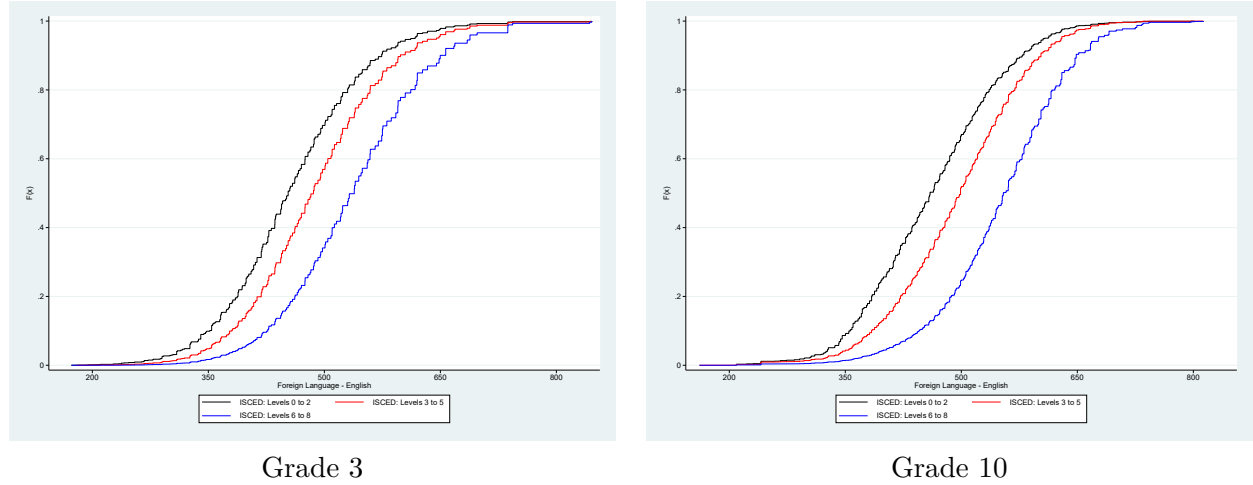


Figure 10: CDFs of English scores according to parents' highest level of education



PARENTAL ENVIRONMENT AND STUDENT ACHIEVEMENT

Table 3: Variables Description

	Type	Values	Description
Students marks			
Mathematics	Num.	0 - 1034	Transformed Mathematics mark
Literature	Num.	0 - 1001	Transformed Literature mark
English (Foreign Language)	Num.	0 - 845	Transformed English mark
Parents' highest education			
Parents' highest education	Qual.	0 - 2	0 ISCED 0 to 2 1 ISCED 3 to 5 2 ISCED 6 to 8
Child's characteristics			
Child's country of birth	Qual.	1 - 2	1 Spain 2 Other
Days/week dedicated to homework	Qual.	1 - 4	1 One day or less 2 Two or 3 days 3 Four or 5 days 4 More than 5
Child's gender	Qual.	1 - 2	1 Female 2 Male
Household's characteristics			
Mother's country of birth	Qual.	1 - 2	1 Spain 2 Other
Father's country of birth	Qual.	1 - 2	1 Spain 2 Other
Freq. books used at home	Qual.	1 - 4	1 Never or almost never 2 One or 2 times/month 3 One or 2 times/week 4 Every or almost every days
Freq. computer used at home	Qual.	1 - 4	1 Never or almost never 2 One or 2 times/month 3 One or 2 times/week 4 Every or almost every days
Freq. internet used at home	Qual.	1 - 4	1 Never or almost never 2 One or 2 times/month 3 One or 2 times/week 4 Every or almost every days
Nb of books at home	Qual.	1 - 5	1 From 0 to 10 2 From 11 to 50 3 From 51 to 100 4 From 101 to 200 5 More than 200
Mother's labour situation	Qual.	1 - 5	1 Full time employee 2 Part time employee 3 Unemployed looking for a job 4 Retired 5 Don't have and don't look for a job
Father's labour situation	Qual.	1 - 5	1 Full time employee 2 Part time employee 3 Unemployed looking for a job 4 Retired 5 Don't have and don't look for a job
Freq. parents talk about school	Qual.	1 - 4	1 Never or almost never 2 One or 2 times/month 3 One or 2 times/week 4 Every or almost every days

PARENTAL ENVIRONMENT AND STUDENT ACHIEVEMENT

Table 3: Variables Description (continued)

	Type	Values	Description
Freq. parents schedule homework	Qual.	1 - 4	1 Never or almost never 2 One or 2 times/month 3 One or 2 times/week 4 Every or almost every days
Freq. parents help for homework	Qual.	1 - 4	1 Never or almost never 2 One or 2 times/month 3 One or 2 times/week 4 Every or almost every days
Freq. parents check homework	Qual.	1 - 4	1 Never or almost never 2 One or 2 times/month 3 One or 2 times/week 4 Every or almost every days
<i>N</i>		327163	

Table 4: Summary Statistics

	Mean	Stand. Dev.
Students marks		
Mathematics	517.4546	95.00513
Literature	505.935	101.8927
English Foreign Language	517.1402	93.7307
Parents' highest education		
Parents highest education	1.494245	.6642053
Child's characteristics		
Child's country of birth	1.044064	.2052369
Days/week dedicated to homework	3.310194	.7581023
Child's gender	1.502942	.4999921
Household's characteristics		
Mother's country of birth	1.165642	.3717597
Father's country of birth	1.159382	.3660328
Freq. books used at home	3.375672	.837836
Freq. computer used at home	3.794766	.5249957
Freq. internet used at home	3.865501	.4511002
Nb of books at home	3.515174	1.203279
Mother's labour situation	1.651739	1.015354
Father's labour situation	1.303231	.7056803
Freq. parents talk about school with child	3.82281	.5012083
Freq. parents schedule homework with child	3.476625	.9059994
Freq. parents help for homework	2.99286	1.098742
Freq. parents check homework with child	3.440297	.9775864
<i>N</i>	327163	

PARENTAL ENVIRONMENT AND STUDENT ACHIEVEMENT

Table 5: Average scores according to grade level

	(1) Gr.3	(2) Gr.6	(3) Gr.10
Mathematics	513.29	521.52	519.33
Literature	514.21	489.35	522.26
English	515.02	515.77	526.17
<i>N</i>	145096	123811	52637

Table 6: Average scores according to grade level, by academic year

	(1) 2016 Gr.3	(2) 2016 Gr.6	(3) 2017 Gr.3	(4) 2017 Gr.6	(5) 2017 Gr.10	(6) 2018 Gr.3	(7) 2018 Gr.6	(8) 2018 Gr.10	(9) 2019 Gr.3	(10) 2019 Gr.6	(11) 2019 Gr.10
Mathematics	507.1	509.8	514.5	515.3	514.1	518.1	545.5	522.9	515.9	518.4	520.9
Literature	510.5	436.3	513.3	514.8	518.7	517.6	524.1	526.5	517.0	516.6	521.1
English	509.2	510.0	515.3	516.1	522.2	520.1	520.1	530.7	517.7	520.5	525.1
<i>N</i>	47061	44249	32390	24578	18029	36535	31060	19594	29110	23924	15014

Table 7: Students distribution according to parent highest education, grade level and academic year

	(1) 2016 Gr.3	(2) 2016 Gr.6	(3) 2017 Gr.3	(4) 2017 Gr.6	(5) 2017 Gr.10	(6) 2018 Gr.3	(7) 2018 Gr.6	(8) 2018 Gr.10	(9) 2019 Gr.3	(10) 2019 Gr.6	(11) 2019 Gr.10
ISCED 0–2	11.15%	10.90%	7.945%	9.520%	9.335%	8.845%	10.15%	8.574%	8.737%	9.296%	8.468%
ISCED 3–5	33.40%	35.43%	29.87%	32.93%	35.53%	28.87%	31.07%	32.08%	26.72%	28.57%	30.14%
ISCED 6–8	55.45%	53.67%	62.19%	57.55%	55.14%	62.28%	58.79%	59.35%	64.54%	62.13%	61.40%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
<i>N</i>	47081	44305	33077	24706	18425	36619	31036	19909	29769	24644	15470

PARENTAL ENVIRONMENT AND STUDENT ACHIEVEMENT

Table 8: Impact of parents' highest level of education – Mathematics

	(1) Mathematics	(2) Mathematics	(3) Mathematics
Parents' highest education			
ISCED 0–2	0 (.)	0 (.)	0 (.)
ISCED 3–5	15.81*** (0.479)	16.72*** (0.517)	10.27*** (0.589)
ISCED 6–8	41.40*** (0.491)	42.92*** (0.530)	29.70*** (0.625)
Constant	482.3*** (0.499)	486.6*** (1.150)	476.5*** (2.661)
Academic year fixed-effect	Yes	Yes	Yes
School fixed-effect	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes
Household's characteristics	No	No	Yes
Observations	415226	367492	316294
Adjusted R^2	0.160	0.172	0.204

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Impact of parents' highest level of education – Literature

	(1) Literature	(2) Literature	(3) Literature
Parents' highest education			
ISCED 0–2	0 (.)	0 (.)	0 (.)
ISCED 3–5	19.25*** (0.515)	20.27*** (0.556)	14.28*** (0.634)
ISCED 6–8	46.42*** (0.527)	48.65*** (0.569)	35.89*** (0.673)
Constant	478.3*** (0.536)	504.4*** (1.237)	488.8*** (2.865)
Academic year fixed-effect	Yes	Yes	Yes
School fixed-effect	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes
Household's characteristics	No	No	Yes
Observations	414805	367757	316529
Adjusted R^2	0.146	0.175	0.197

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Impact of parents' highest level of education – English

	(1) English	(2) English	(3) English
Parents' highest education			
ISCED 0–2	0 (.)	0 (.)	0 (.)
ISCED 3–5	21.81*** (0.437)	21.75*** (0.471)	16.06*** (0.535)
ISCED 6–8	52.55*** (0.448)	52.34*** (0.483)	39.68*** (0.568)
Constant	473.7*** (0.455)	492.3*** (1.048)	468.3*** (2.418)
Academic year fixed-effect	Yes	Yes	Yes
School fixed-effect	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes
Household's characteristics	No	No	Yes
Observations	413466	365774	314863
Adjusted R^2	0.293	0.303	0.329

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Impact of parents' highest level of education according to grade level – **Grade 3 vs Grade 6**

	Mathematics			Literature			English		
	(1) Math.	(2) Math.	(3) Math.	(1) Lit.	(2) Lit.	(3) Lit.	(1) Engl.	(2) Engl.	(3) Engl.
Parents highest education									
ISCED 0–2	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
ISCED 3–5	18.84*** (0.757)	19.05*** (0.762)	12.05*** (0.864)	19.91*** (0.845)	20.39*** (0.840)	13.30*** (0.950)	18.59*** (0.698)	19.22*** (0.700)	13.35*** (0.794)
ISCED 6–8	46.16*** (0.745)	46.16*** (0.750)	30.22*** (0.874)	44.50*** (0.832)	45.18*** (0.826)	28.44*** (0.960)	46.57*** (0.687)	47.07*** (0.689)	33.11*** (0.803)
Grade 3 vs Grade 6									
I(3,6)	9.017*** (0.932)	10.49*** (0.940)	-2.856*** (1.081)	-30.56*** (1.040)	-29.72*** (1.035)	-42.57*** (1.188)	-3.799*** (0.859)	-3.022*** (0.862)	-13.54*** (0.993)
Interactions									
ISCED 0–2 × I(3,6)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
ISCED 3–5 × I(3,6)	-0.759 (1.083)	-1.029 (1.090)	-1.161 (1.219)	1.955 (1.210)	2.237* (1.201)	2.442* (1.339)	4.857*** (0.998)	4.882*** (1.000)	5.017*** (1.119)
ISCED 6–8 × I(3,6)	-0.0122 (1.021)	0.0649 (1.027)	0.626 (1.155)	10.85*** (1.140)	11.26*** (1.133)	12.21*** (1.269)	8.692*** (0.940)	9.089*** (0.943)	9.549*** (1.060)
Constant									
	475.5*** (0.740)	483.6*** (1.329)	485.8*** (3.166)	491.3*** (0.826)	518.2*** (1.467)	508.6*** (3.479)	477.8*** (0.682)	498.6*** (1.222)	483.3*** (2.908)
Academic year Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Household's characteristic	No	No	Yes	No	No	Yes	No	No	Yes
Observations	310508	306732	264386	311019	307233	264811	308884	305117	263032
Adjusted R^2	0.178	0.186	0.222	0.162	0.192	0.224	0.301	0.312	0.337

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ $I(u, v)$ is a dummy variable which takes the value 0 if $g = u$ (reference grade), and 1 if $g = v$

Table 12: Impact of parents' highest level of education according to grade level – **Grade 6 vs Grade 10**

	Mathematics			Literature			English		
	(4) Math.	(5) Math.	(6) Math.	(4) Lit.	(5) Lit.	(6) Lit.	(4) Engl.	(5) Engl.	(6) Engl.
Parents highest education									
ISCED 0–2	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
ISCED 3–5	18.40*** (0.819)	17.86*** (0.816)	11.03*** (0.921)	21.82*** (0.877)	22.05*** (0.884)	14.96*** (0.993)	22.10*** (0.727)	22.29*** (0.719)	15.87*** (0.806)
ISCED 6–8	47.02*** (0.833)	46.62*** (0.833)	31.28*** (0.966)	52.57*** (0.892)	53.50*** (0.902)	36.83*** (1.042)	51.79*** (0.739)	52.93*** (0.734)	37.69*** (0.846)
Grade 6 vs Grade 10									
I(6,10)	6.615*** (1.205)	13.98*** (1.506)	-8.327*** (1.696)	15.05*** (1.297)	21.71*** (1.637)	3.264* (1.837)	-2.601** (1.069)	3.110** (1.326)	-10.94*** (1.484)
Interactions									
ISCED 0–2 × I(6,10)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
ISCED 3–5 × I(6,10)	-9.866*** (1.229)	-11.76*** (1.545)	-7.850*** (1.714)	-5.898*** (1.324)	-7.829*** (1.680)	-6.367*** (1.858)	1.979* (1.090)	1.604 (1.359)	2.445 (1.499)
ISCED 6–8 × I(6,10)	-19.15*** (1.220)	-20.84*** (1.529)	-13.61*** (1.700)	-15.16*** (1.314)	-17.01*** (1.662)	-11.55*** (1.842)	7.618*** (1.083)	6.814*** (1.346)	9.885*** (1.488)
Constant									
	474.7*** (0.822)	469.3*** (1.784)	464.1*** (4.107)	395.9*** (0.879)	409.4*** (1.939)	390.1*** (4.434)	474.3*** (0.729)	483.2*** (1.572)	457.9*** (3.599)
Academic year Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Household's characteristic	No	No	Yes	No	No	Yes	No	No	Yes
Observations	248339	202678	173619	247527	202556	173528	247798	202173	173209
Adjusted R^2	0.155	0.167	0.206	0.209	0.249	0.282	0.315	0.328	0.360

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ $I(u, v)$ is a dummy variable which takes the value 0 if $g = u$ (reference grade), and 1 if $g = v$

Table 13: Impact of parents' highest level of education according to grade level – **Grade 3 vs Grade 10**

	Mathematics			Literature			English		
	(7) Math.	(8) Math.	(9) Math.	(7) Lit.	(8) Lit.	(9) Lit.	(7) Engl.	(8) Engl.	(9) Engl.
Parents' highest education									
ISCED 0–2	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
ISCED 3–5	18.97*** (0.769)	19.00*** (0.765)	12.33*** (0.877)	21.19*** (0.751)	21.90*** (0.745)	15.32*** (0.850)	19.19*** (0.711)	19.93*** (0.708)	14.05*** (0.806)
ISCED 6–8	45.24*** (0.777)	45.39*** (0.776)	30.08*** (0.917)	46.60*** (0.758)	48.01*** (0.754)	32.75*** (0.888)	45.51*** (0.718)	47.29*** (0.717)	33.20*** (0.843)
Grade 3 vs Grade 10									
I(3,10)	9.617*** (1.163)	17.89*** (1.463)	-12.96*** (1.707)	9.118*** (1.140)	16.65*** (1.428)	-8.665*** (1.661)	-11.74*** (1.072)	-5.893*** (1.348)	-28.24*** (1.565)
Interactions									
ISCED 0–2 $\times$ I(3,10)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
ISCED 3–5 $\times$ I(3,10)	-10.50*** (1.187)	-12.24*** (1.500)	-8.859*** (1.684)	-5.154*** (1.165)	-7.210*** (1.466)	-6.144*** (1.639)	4.794*** (1.094)	4.653*** (1.382)	5.232*** (1.543)
ISCED 6–8 $\times$ I(3,10)	-16.23*** (1.175)	-16.64*** (1.484)	-9.076*** (1.675)	-7.959*** (1.152)	-9.132*** (1.449)	-3.856** (1.630)	16.01*** (1.083)	16.55*** (1.368)	19.72*** (1.535)
Constant									
	474.8*** (0.769)	482.1*** (1.410)	480.3*** (3.225)	475.5*** (0.750)	499.9*** (1.373)	488.1*** (3.129)	478.4*** (0.712)	497.2*** (1.304)	478.7*** (2.961)
Academic year Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Household's characteristic	No	No	Yes	No	No	Yes	No	No	Yes
Observations	271605	225574	194583	271064	225725	194719	270250	224258	193485
Adjusted R^2	0.180	0.200	0.225	0.152	0.176	0.205	0.302	0.314	0.337

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ $I(u, v)$ is a dummy variable which takes the value 0 if $g = u$ (reference grade), and 1 if $g = v$

Table 14: Impact of parental investment and student's effort

	Mathematics			Literature			English		
	(1) Gr.3	(2) Gr.6	(3) Gr.10	(4) Gr.3	(5) Gr.6	(6) Gr.10	(7) Gr.3	(8) Gr.6	(9) Gr.10
Parents' highest education									
ISCED 0–2	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
ISCED 3–5	12.04*** (0.861)	10.99*** (0.909)	3.304** (1.496)	15.27*** (0.864)	14.68*** (1.054)	7.688*** (1.325)	14.71*** (0.806)	16.40*** (0.806)	16.46*** (1.290)
ISCED 6–8	30.01*** (0.914)	30.84*** (0.972)	17.90*** (1.584)	32.99*** (0.917)	35.96*** (1.128)	24.54*** (1.403)	35.62*** (0.855)	39.08*** (0.862)	42.99*** (1.366)
Days/week dedicated to homework									
One day or less	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2 or 3 days	-0.185 (1.533)	2.338 (2.288)	9.076*** (2.978)	-2.184 (1.538)	7.045*** (2.658)	9.176*** (2.648)	-0.957 (1.437)	3.693* (2.030)	3.954 (2.572)
4 or 5 days	-3.599** (1.516)	5.090** (2.194)	10.51*** (2.863)	-5.920*** (1.521)	10.56*** (2.548)	14.59*** (2.545)	-5.995*** (1.421)	4.559** (1.947)	7.796*** (2.473)
More than 5	-10.90*** (1.548)	4.527** (2.203)	14.05*** (2.866)	-15.81*** (1.553)	9.282*** (2.558)	19.23*** (2.548)	-14.08*** (1.451)	3.986** (1.955)	12.68*** (2.476)
Freq. parents talk about school									
Never or almost never	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
1 or 2 times/month	5.461 (4.141)	-0.716 (3.151)	4.868* (2.705)	2.225 (4.152)	-1.567 (3.661)	-0.793 (2.394)	0.894 (3.862)	-0.611 (2.798)	3.610 (2.339)
1 or 2 times/week	9.676*** (3.545)	2.266 (2.650)	6.345*** (2.303)	10.45*** (3.553)	7.938*** (3.077)	4.034** (2.038)	4.841 (3.299)	4.593* (2.353)	9.155*** (1.992)
Every or almost every day	10.57*** (3.485)	4.455* (2.585)	9.023*** (2.228)	16.06*** (3.492)	14.41*** (3.001)	8.285*** (1.970)	8.704*** (3.243)	9.119*** (2.295)	12.91*** (1.927)
Constant	482.3*** (4.572)	473.2*** (4.881)	470.8*** (9.367)	491.0*** (4.582)	472.2*** (5.657)	463.2*** (8.268)	483.7*** (4.265)	463.5*** (4.331)	441.5*** (8.148)
Academic year Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household's characteristic	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	142675	121711	51908	143001	121810	51718	141654	121378	51831
Adjusted R^2	0.262	0.235	0.202	0.230	0.301	0.200	0.361	0.378	0.343

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Impact of parent's highest level of education (predicted values by first stage IV regression) according to grade level - **Grade 3 vs Grade 6**

	Mathematics			Literature			English		
	(1) Math.	(2) Math.	(3) Math.	(1) Lit.	(2) Lit.	(3) Lit.	(1) Engl.	(2) Engl.	(3) Engl.
Parent's highest education	67.52*** (1.062)	65.63*** (1.058)	46.97*** (11.74)	63.02*** (1.171)	65.58*** (1.151)	59.24*** (12.82)	57.75*** (0.985)	58.80*** (0.976)	69.05*** (10.85)
Grade 3 vs Grade 6 I(3,6)	9.682*** (2.262)	8.517*** (2.251)	-8.106*** (2.239)	-36.42*** (2.494)	-36.62*** (2.449)	-53.68*** (2.442)	-15.35*** (2.097)	-16.14*** (2.076)	-28.49*** (2.070)
Interactions									
	-1.522 (1.477)	0.272 (1.470)	2.828* (1.457)	10.11*** (1.628)	10.97*** (1.600)	14.26*** (1.588)	11.00*** (1.369)	12.24*** (1.356)	13.88*** (1.346)
Constant	409.4*** (1.662)	425.6*** (2.075)	466.5*** (8.471)	429.1*** (1.834)	457.3*** (2.259)	484.7*** (9.246)	427.6*** (1.542)	451.4*** (1.916)	453.3*** (7.824)
Academic year Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Household's characteristics	No	No	Yes	No	No	Yes	No	No	Yes
Observations	248466	248466	248466	248816	248816	248816	247207	247207	247207
Adjusted R^2	0.171	0.180	0.215	0.146	0.177	0.209	0.278	0.293	0.321

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I(u, v) is a dummy variable which takes the value 0 if $g=u$ (reference grade), and 1 if $g=v$

Table 16: Impact of parent's highest level of education (predicted values by first stage IV regression) according to grade level - **Grade 6 vs Grade 10**

	Mathematics			Literature			English		
	(4) Math.	(5) Math.	(6) Math.	(4) Lit.	(5) Lit.	(6) Lit.	(4) Engl.	(5) Engl.	(6) Engl.
Parent's highest education	66.80*** (1.212)	66.13*** (1.206)	34.14** (13.74)	70.74*** (1.305)	73.56*** (1.284)	40.30*** (14.66)	65.08*** (1.078)	66.82*** (1.069)	76.60*** (12.14)
Grade 6 vs Grade 10 I(6,10)	22.26*** (3.380)	25.73*** (3.370)	1.693 (3.384)	35.62*** (3.649)	37.89*** (3.598)	16.24*** (3.617)	-10.08*** (3.005)	-8.706*** (2.986)	-29.77*** (2.993)
Interactions									
	-18.06*** (2.107)	-19.85*** (2.101)	-14.54*** (2.099)	-18.64*** (2.274)	-19.89*** (2.243)	-15.37*** (2.243)	10.71*** (1.872)	9.912*** (1.861)	15.44*** (1.856)
Constant	411.4*** (1.905)	407.8*** (2.606)	447.5*** (9.320)	333.5*** (2.051)	344.3*** (2.780)	375.1*** (9.948)	417.8*** (1.694)	426.0*** (2.309)	422.8*** (8.241)
Academic year Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Household's characteristics	No	No	Yes	No	No	Yes	No	No	Yes
Observations	166358	166358	166358	166220	166220	166220	165923	165923	165923
Adjusted R^2	0.149	0.159	0.197	0.197	0.223	0.257	0.291	0.304	0.338

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I(u, v) is a dummy variable which takes the value 0 if $g=u$ (reference grade), and 1 if $g=v$

Table 17: Impact of parent's highest level of education (predicted values by first stage IV regression) according to grade level - **Grade 3 vs Grade 10**

	Mathematics			Literature			English		
	(7) Math.	(8) Math.	(9) Math.	(7) Lit.	(8) Lit.	(9) Lit.	(7) Engl.	(8) Engl.	(9) Engl.
Parent's highest education	65.68*** (1.101)	63.94*** (1.097)	36.96*** (12.70)	63.71*** (1.076)	65.60*** (1.064)	48.73*** (12.33)	57.45*** (1.022)	58.61*** (1.016)	50.93*** (11.72)
Grade 3 vs Grade 10 I(3,10)	19.91*** (3.243)	20.19*** (3.233)	-16.85*** (3.397)	18.21*** (3.178)	17.95*** (3.145)	-11.83*** (3.304)	-34.30*** (3.005)	-35.37*** (2.991)	-67.77*** (3.136)
Interactions									
	-14.48*** (2.026)	-12.74*** (2.023)	-4.696** (2.094)	-9.742*** (1.986)	-8.109*** (1.968)	-2.643 (2.036)	23.12*** (1.878)	24.98*** (1.872)	32.48*** (1.933)
Constant									
	413.9*** (1.719)	425.9*** (2.170)	467.7*** (8.311)	418.3*** (1.679)	441.9*** (2.108)	469.1*** (8.067)	429.7*** (1.595)	449.4*** (2.012)	466.6*** (7.670)
Academic year Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed-Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child's characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Household's characteristics	No	No	Yes	No	No	Yes	No	No	Yes
Observations	186008	186008	186008	186130	186130	186130	185028	185028	185028
Adjusted R^2	0.183	0.193	0.218	0.149	0.171	0.197	0.284	0.294	0.320

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I(u, v) is a dummy variable which takes the value 0 if $g=u$ (reference grade), and 1 if $g=v$