CENTRE FOR HOUSEHOLD, INCOME, LABOUR AND DEMOGRAPHIC ECONOMICS

Do Boys and Girls Perform Better at Math Just Studying More ?

Eleonora Matteazzi, Martina Menon and Federico Perali

No. 96

December 2021

CHILD WORKING PAPERS

www.carloalberto.org/child-collegiocarloalberto/working-papers

UNIVERSITÀ DEGLI STUDI DI TORINO

Do Boys and Girls Perform Better at Math Just Studying More?

Eleonora Matteazzi*, Martina Menon†, and Federico Perali‡

Abstract

This paper investigates the role of effort on mathematics performance of boys and girls, an aspect that may contribute to our understanding of the gender gap in science, technology, engineering and math (STEM) fields in college. We exploit a remarkably rich primary data set to estimate a simultaneous equations model of mathematics attainment and students' effort. Our estimation strategy infers causal relations by relying on an instrumental variable approach validated using weak-instruments-robust confidence sets and partial identification techniques. The results show that study effort plays a different role in the math performance of girls and boys. If a boy dedicates one extra hour to study, his math grade increases by 1 point on a 10-point scale. Differently, an additional hour of home study does not have an effect on girls' math performance, though, in our sample, on average, girls perform significantly better than boys in math. We also examine the role played by peers, the quality of the attended school, and family socio-economic background. These factors mainly affect math achievement only indirectly through student's effort. Validity tests suggest that our results are not confounded by unobservable heterogeneity. Our findings suggest that asking girls for additional efforts may not be effective to bridge the gender gap in STEM.

Keywords: Mathematics, effort, gender inequality, peer effects, school quality.
JEL Classification: I21, I28.
Funding: This work was supported by the European Investment Bank (EIB) Institute within the "STAges de REcherche BEI-EIB" research programme (STAREBEI 2008).

*Department of Economics and Management, University of Verona, Via Cantarane 24, 37129 Verona (Italy), and EconomiX-CNRS, Avenue de la République 200, 92001 Nanterre cedex (France); eleonora. matteazzi@univr.it.

†Department of Economics and Management, University of Verona, Via Cantarane 24, 37129 Verona (Italy). E-mail: martina.menon@univr.it.

‡Department of Economics and Management, University of Verona, Via Cantarane 24, 37129 Verona (Italy), and Center for Household, Income, Labor and Demographics (CHILD). E-mail: federico.perali@univr.it.

1 Introduction

The main objective of this study is to investigate math performance of high school leavers and the extent to which studying more results in a similar or different effect on the math achievement of male and female students. The effectiveness in studying math and the acquisition of math skills in high schools significantly influences the choice to enroll in STEM graduate majors (Speer 2017, Delaney and Devereux 2019, 2021, Card and Payne 2021) and pursue STEM careers, where women are underrepresented. According to the OECD statistics for the year 2019, in Europe only 16% of women enrolled in tertiary education have undertaken a STEM field of study, against 42% of men. The gender gap is also particularly high in some of the fastest-growing and highest-paid jobs of the future that are especially relevant for the post-pandemic restart of the global economy (Di Tommaso *et al.* 2021). Hence, the study of how effort impacts on the acquisition of math skills is relevant to understand the gender gap in mathematics and in STEM subjects in universities.

Effort is one of the most important inputs of the educational production function. Students are normally held responsible both for the level of effort put into studying and the quality of concentration exerted during their study time. Educational attainment is also influenced by environmental circumstances, which include aspects related to childhood and family environment, as well as genetic variables, for which the student is not responsible (Roemer and Trannoy 2016). Parents play a crucial role in the transmission of cognitive and non-cognitive abilities to their children (Carneiro and Heckman 2003, Cuhna *et al.* 2006) and they can also affect their children's decision about how much effort exert through supervision or providing appropriate incentives. School and teachers quality and the influence of peers are also relevant for school achievement (Wößmann *et al.* 2005, Sacerdote 2011, Mehta *et al.* 2018).

Empirical analyses on the role of effort on schooling outcomes are scarce and often do not reach clear conclusions mainly because data with reliable measures of students' effort are limited. Among the existing literature, few empirical works employ data with the information on hours spent doing homework or studying to infer students' effort (Stinebrickner and Stinebrickner 2008, Eren and Henderson 2011, Kuehn and Landeras 2014). Other studies have adopted qualitative indicators of effort such as students' attitudes towards school attendance or homework, or students' energy investments to be successful in studies (De Fraja *et al.* 2010). Kuehn and Landeras (2014) investigate the effect of hours of homework per week on a standardized test taken by sixth and ninth graders of the Madrid region. Eren and Henderson (2011) examine the effect of homework on math, science, English and history test scores for eighth grade students in the United States. On the other hand, estimates of the causal effect of effort on educational outcomes are scant. Stinebrickner and Stinebrickner (2008) evaluate the effect of study hours on the grade point average in the first semester of students of the Berea College in Kentucky. De Fraja *et al.* (2010) analyze the effect of students' attitudes, such as laziness as perceived by the teacher and students' perception about school being a "waste of time," on examination results using the 1958 National Child Development Study.

The present work contributes to this strand of literature by proposing a causal study on the effect of study effort on schooling achievements accounting for gender differences. Our estimation strategy infers causal relations by relying on an instrumental variable approach validated using weak-instruments-robust confidence sets (Andrews 2018, Andrews *et al.* 2019) and partial identification techniques (Kiviet 2020). Unlike previous work, our measure of schooling outcomes are math school grades and not standardized test scores such as PISA (Jansen *et al.* 2016). In general, grades measure subject knowledge and skills, besides reflecting students' effort, motivation, and communicative abilities (McMillan *et al.* 2002), which in turn may affect the way in which teachers assign grades. However, teachers of mathematics assign grades in a "right or wrong" way and their grading practice is more objective (Pilcher 1994). The empirical analysis exploits a uniquely rich primary data set collected by the authors providing detailed information about students' actual and past schooling outcomes and a quantitative measure of effort specified as the number of hours spent in studying in a typical school day. Information on family background, students' risk and time preferences, their future working ambitions are also gathered along with the students' intentions to pursue further education after high school. The questionnaire design allows us to derive several measures of peer effects. Data are also complemented with external information about school quality.

The analysis of whether effort can improve math performance is likely affected by omitted variable bias because there can be unobservable confounding factors, such as innate aptitudes of the student, correlated with both math performance and effort. Given the complexity of implementing randomized trial designs to exogenously change students' effort, we adopt an instrumental variable method to identify the causal effect of effort on schooling outcome using as instruments the gender-specific local unemployment rate and the average local daily rainfall intensity in the school year. Current local labour market conditions may impact students' expectation about future employment prospects. In turn, this subjective expectation can affect current effort provision and, indirectly, schooling outcome (Chadi *et al.* 2019). Effort provision may also be affected by weather conditions. In the literature, there is evidence that weather has a significant impact on mood, cognition and productivity (Keller *et al.* 2005, Simonsohn 2010, Lee *et al.* 2014). We expect that raining days reduce the opportunity cost of out-of-home activities and foster both the quantity and quality of study effort.

From the perspective of the novelty of the results, the study finds that effort plays a significant role in the production of math outcome for boys only. Our estimates show that an additional hour devoted to study raises boys' math grade by 1 point on a 10-point scale. Girls are more diligent than boys, but an additional hour of out-of-school study does not improve their math scores. These results suggest that while for boys it is effective to provide incentives to study harder to improve their math performance, for girls it is probably more constructive to implement programs that make girls less susceptible to the negative effects of math anxiety and sterotypes and reinforce their work aspirations for STEM occupations. In the literature, there is wide evidence that math anxiety has a negative impact on performance and acquisition of math skills and that it is much more common among women than men, at all life-cycle stages (Van Mier *et al.* 2019). It has also been shown that women are less confident than men in their abilities (Bian *et al.* 2017, Di Tommaso *et al.* 2021, Jiang 2021), a factor that can push them to study more hours, without necessarily getting better grades in math. Our study finds that boys and girls behave significantly different also when facing bad job market prospects. Girls are more likely to reduce their effort provision when labour market conditions worsen. On the other hand, boys would exert more effort in presence of bad employment outlooks. In addition, rain precipitation shows opposite effects across gender, with boys more likely than girls to allocate more hours to study activities in case of raining days. Further, our findings show that peers' characteristics and school quality play a more important role than family background variables and that they are more likely to affect student effort than having a direct association with math outcome. The validity tests of our IV estimates based on Kiviet (2020) suggest that our results are not confounded by unobservable heterogeneity.

From a policy perspective, a better understanding of how students' effort decisions impact on math achievement provide useful insights to understand the gender gap in mathematics and the under-representation of women in STEM fields in universities. Educational programs tailored to high school students should aim at reducing gender disparities in the production of mathematical skills and reducing socio-cultural and psychological barriers that limit women's interests, preferences and participation in many careers, especially those in the STEM fields. In our sample, among high school leavers who want to pursue further education, only 8% of girls intend to undertake a STEM major, against 34% of boys. Similar figures are observed also among students enrolled in scientific high schools and top-performing math students. About 20% of both girls and boys receive a math grade equal or greater than 8 out of 10, but only 15% of girls as opposed to 42% of boys are willing to undertake a STEM major.

The plan of the paper is as follows. Section 2 presents the theoretical framework explaining students' behaviour in terms of schooling attainment and effort provision at the basis of the specification of the econometric model. Section 3 describes the survey design and the main features of the primary data. Section 4 presents the empirical methods. Section 5 shows the estimates of the education production function and validity tests of our instrumental variables along with the robustness analysis. Empirical results of a sensitivity analysis are also presented. Section 6 concludes and discusses policy implications.

2 The Educational Production Function and Effort

We explain the behaviour of students in terms of schooling attainment and effort provision. Schooling choices are represented using a two period model, $t = 1, 2$. In period $t = 1$, the student is enrolled in education, where she chooses her optimal level of effort E^1, which is expected to impact her future employment outlook and earnings through its effect on schooling achievements. In period $t = 2$, she is expected to work and earn income. In each period, student preferences over leisure L^t and income I^t are represented by a sub-utility function U^t increasing and concave in its arguments. The student intertemporal utility function is then defined as

$$V = U^1\left(L^1, I^1\right) + \beta U^2\left(L^2, I^2\right) \qquad \text{with} \qquad \beta > 0, \tag{1}$$

where L^1 $\left(L^2\right)$ is leisure consumption in $t = 1$ $(t = 2)$ and I^1 $\left(I^2\right)$ is income in $t = 1$ $(t = 2)$. As in Chadi *et al.* (2019) L^2 and I^1 are considered exogenous. β is the subjective discount factor.[1] In $t = 1$, the student has to choose how much effort to put into studying. The individual time constraint implies that

$$T^1 = L^1 + E^1, \tag{2}$$

where E^1 is the self-study time, such as the number of hours allocated to study at home, and T^1 is the total amount of time available to the student. E^1 is an indicator of effort.[2]

The educational production function is described by the following equation

$$Y^1 = f(E^1, X^1, P^1, S^1, \sigma), \tag{3}$$

where the schooling outcome Y^1 is assumed to be a function of effort E^1, a set of observable individual and family characteristics X^1, peers' attributes P^1, school quality S^1, and other time invariant exogenous factors σ, mainly unobservables, such as individual ability or innate talent. We assume the EPF being increasing in effort but at a decreasing rate, $f_{E^1} > 0, f_{E^1E^1} < 0$.

In $t = 2$, the expected income I^2 is linked to schooling attainment as

$$I^2 = \mu f(E^1, X^1, P^1, S^1, \sigma), \tag{4}$$

where $\mu > 0$ is the student's job market expectation (Chadi *et al.* 2019). In presence of bad job market expectations (low values of μ), a given schooling attainment will be associated with a lower expected income, either because of the possibility of experiencing fragmented employment paths, which may cause some income insecurity, or because of low earnings prospects in presence of high unemployment. On the other hand, if job prospects are good (high values of μ), a given level of schooling outcome will entail a higher expected income. In Chadi *et al.* (2019) μ is a measure for non-ability-related labor market expectations, it is exogenously given and is independent of ability and effort decisions.[3]

Using equations (2), (3) and (4), we can rewrite the student's inter-temporal utility function as

$$V = U^1\left(T^1 - E^1, I^1\right) + \beta U^2\left(L^2, \mu f\left(E^1, X^1, P^1, S^1, \sigma\right)\right). \tag{5}$$

In $t = 1$, the student decides about the amount of hours she will spend studying, while the remaining time is allocated to leisure. Maximisation of the inter-temporal utility function (5)

[1] To keep our analysis as simple as possible, the discount rate is considered as exogenously given and independent of students' characteristics (De Fraja *et al.* 2010, Chadi *et al.* 2019).

[2] One can distinguish between a qualitative and a quantitative dimension of individual effort. While the quantitative dimension is measured by the study time, the qualitative element regards how concentrated the student is during the study time. We focus on the quantitative dimension of effort.

[3] In the empirical application the authors use the local unemployment rate to instrument students' subjective perceptions of job market prospects, because they argue that the latter can be related to student's ability and therefore it would be impossible to disentangle job market prospects from individual ability.

leads to the following first order condition

$$-U^1_{E^1} + \beta U^2_{I^2} \mu f_{E^1} = 0, \tag{6}$$

where $-U^1_{E^1}$ is the marginal cost of a one-hour increase in self-study, that is study entails a utility cost, and $\beta U^2_{I^2} \mu f_{E^1}$ is the marginal gain of an additional hour of effort in terms of future expected income.[4] Equation (6) says that a utility reduction in $t = 1$, due to an increased amount of study hours and a reduced consumption of leisure, is perfectly compensated by an increased utility in $t = 2$, due to a larger expected income because of an improved schooling attainment. From equation (6) we get the provision of optimal effort $E^{1*} = E^1(X^1, P^1, S^1, \mu, \beta, \sigma)$. Substituting E^{1*} into the EPF, we derive the optimal value of schooling outcome $Y^* = f(E^{1*}, X^1, P^1, S^1, \sigma) = \bar{f}\left(X^1, P^1, S^1, \mu, \beta, \sigma\right)$. Our objective is the relationship between effort and schooling outcome and will be investigated below.

Totally differentiating equation (6) provides the following relationship

$$\frac{dE^1}{d\mu} = -\frac{U^1_{E^1\mu} + \beta U^2_{I_2} f_{E^1} + \beta U^2_{I_2 I_2} \mu f_{E^1\mu}}{U^1_{E^1 E^1} + \beta U^2_{I_2 I_2} \mu f_{E^1 E^1}} \lesseqgtr 0, \tag{7}$$

where $U^1_{E^1 E^1} < 0$, $U^2_{I_2} > 0$, $U^2_{I_2 I_2} < 0$, $f_{E^1} > 0$, and $f_{E^1 E^1} < 0$, while the sign of the derivatives $U^1_{E^1\mu}$ and $f_{E^1\mu}$ are *a priori* unknown, producing the ambiguity of the marginal effect $\frac{dE^1}{d\mu}$. The sign of expression (7) will be investigated in the empirical analysis, where μ is measured by the local unemployment rate.

Suppose now that labour market prospects worsen, that is μ decreases. This can be due, for example, to a rise in unemployment risk. Two possible scenarios may exist. The first one is the case of a *discouraged* student. Confronted with bad job market prospects and difficulties in entering employment after school completion, a discouraged student is expected to reduce her effort provision believing that the achievement of a better schooling outcome will not be enough to succeed in the labour market. For this type of student, we expect $dE^1/d\mu > 0$.[5] Notice that, the decision to reduce the effort provision in $t = 1$ has as a consequence the achievement of a worse schooling outcome. This, together with bad job market prospects, results in a much lower expected income in $t = 2$. The second scenario is the case of a *motivated* student. This type of student is expected to put as much effort as possible in $t = 1$ to improve her chances of having a successful schooling outcome, which can reduce obstacles to labour market entry in $t = 2$. Notice that, the improved schooling achievement (due to the higher effort put into studying in $t = 1$) could result in a lower reduction of expected income in $t = 2$. For a motivated student we expect $dE^1/d\mu < 0$. These two possible scenarios are represented in Figure 1. Panel *a* refers to a *discouraged* student, while panel *b* to a *motivated* one.

In Figure 1, continuous lines refer to the initial situation, while dashed lines are the result of a worsening in job market prospects (diminishing μ). The equilibrium moves from the initial point A° to the final point $A^{'}$. Notice that for a *discouraged* student (Panel *a*), a worsening in

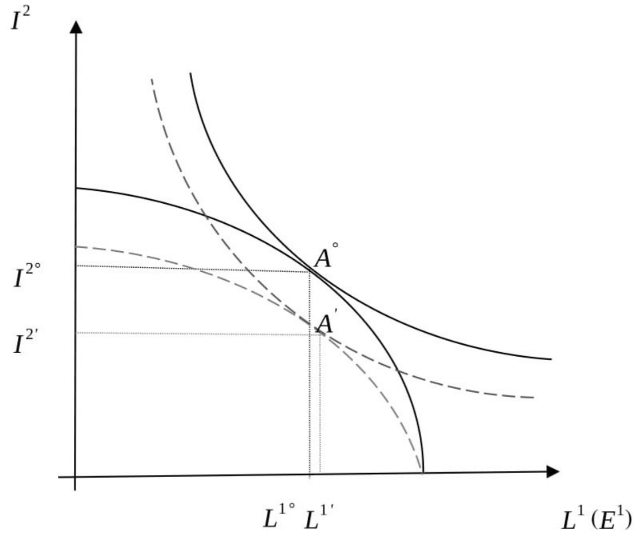
[4]Sub-scripts denote partial derivatives.

[5]For discouraged students, a negative change in μ causes a negative change in E^1 leading the total differentiation $dE^1/d\mu$ to be positive.

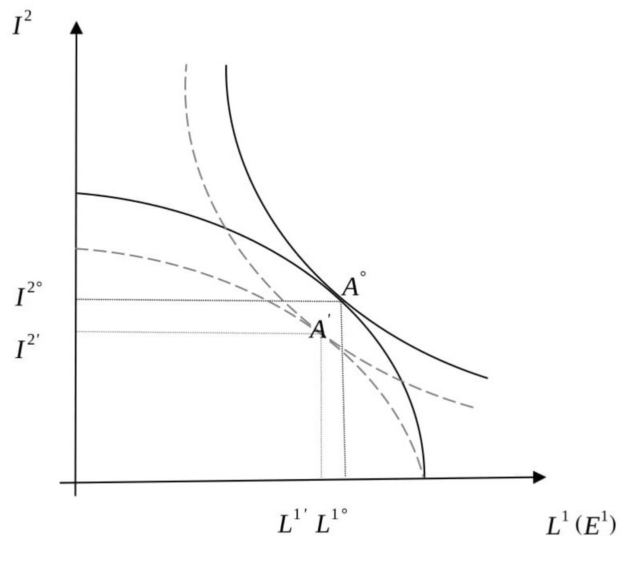
labour market opportunities is associated with a reduction in effort provision (increase in leisure consumption) and a reduction in expected income in $t = 2$. For a *motivated* student (Panel b), it is associated with an increase in effort provision (decrease in leisure consumption), but a lower reduction in future expected income.

Figure 1: Comparative Statics

Panel *a*. *Discouraged* students



Panel *b*. *Motivated* students



Notes: I^2 denotes income in $t = 2$, L^1 leisure consumption in $t = 1$, and E^1 the number of hours allocated to study at home. Continuous lines refer to the initial situation. Dashed lines are the result of a worsening in job market prospects. A° is the initial equilibrium, while A' is the equilibium resulting from the worsening in job market prospects.

As previously argued, students' decision about how much hours allocate to study involves an inter-temporal trade-off: by forgoing leisure and increasing study effort today, the student bears a cost immediately, which will result in a gain tomorrow. We can distinguish two types of students: the *forward-looking* and the *short-sighted* ones. *Forward-looking* students, being more patient, are expected to exert more effort in $t = 1$, resulting in higher educational outcome and performance, which would raise their chances to earn a higher income in $t = 2$. In contrast, *short-sighted* students, connoted by less patience, would prefer to enjoy more leisure time today, resulting in lower educational attainment and, possibly, lower earnings in the future. Hence, *forward-looking* students are more patient than *short-sighted* students. Totally differentiating equation (6) provides the following relationship

$$\frac{dE^1}{d\beta} = -\frac{U^1_{E^1\beta} + U^2_{I_2}\mu f_{E^1} + \beta U^2_{I_2I_2}\mu f_{E^1\beta}}{U^1_{E^1E^1} + \beta U^2_{I_2I_2}\mu f_{E^1E^1}}, \tag{8}$$

where $U^1_{E^1E^1} < 0$, $U^2_{I_2} > 0$, $U^2_{I_2I_2} < 0$, $f_{E^1} > 0$, and $f_{E^1E^1} < 0$, while the sign of the derivatives $U^1_{E^1\beta}$ and $Y_{E^1\beta}$ are *a priori* unknown. However, given that the lower the beta, the lower the degree of patience of an individual, we should expect that $dE^1/d\beta > 0$. On the other hand, the higher the β, the higher the degree of patience and, hence, the higher the provision of student effort (Segal 2013, Non and Tempelaar 2016).

3 Survey Design and Data

Our data come from a unique school-based survey conducted by the authors from March to April 2009 aimed at investigating the educational choice, effort provision and schooling attainments of high school leavers[6] living in Veneto, a region located in the North-East of Italy. The decision to focus on this Italian region was based mainly on the peculiar links between the educational environment and the economic performance of the region. The INVALSI (2009) has found that students in the North-East of Italy, especially in the Veneto region, are the highest achievers in the country.

The survey is representative of the 466 high schools of the Veneto region, classifying the educational institutions on a geographical basis and allowing stratification of the sample according to the 579 municipalities and seven provinces of the region.[7] High schools were preliminarily contacted by e-mail and by telephone to illustrate the research objectives and its relevance for students. Sampling was checked daily to achieve a proportional stratified sample. When a stratum was not proportional to its corresponding size in the population, the high schools belonging to that stratum were contacted again through a follow-up recall session to obtain the most balanced sample possible considering the voluntary basis of the participation.

Data were collected using a computer assisted self-interview survey method (CASI) held in

[6]Italian high school consists of five years of study, which lead to a diploma issued by the State upon successful completion of national exams. Within the high school structure, the most challenging level of instruction and studies are offered at lyceums, which specifically prepare students for university. Lyceum studies are offered in a number of disciplines, such as classics literature, math, science, technology, languages, and fine and performing arts. The "standard" track of high school leads to university or to professional training at the post-secondary level. Italy offers technical schools for agricultural, industrial, and commercial sectors, while vocational schools lead to specific trades. Moreover, vocational schools do not necessarily foresee university study and may lead directly to work after secondary education.

[7]Veneto provinces are Belluno, Padova, Rovigo, Treviso, Venezia, Verona, and Vicenza.

the school's computing rooms acting as our "lab-in-the-field" by providing access to a dedicated website (Levitt and List 2007, 2009, List 2009). The questionnaire underwent a pre-test period of two weeks and was programmed using SNAP to prompt screen messages when responses where inconsistent. A total of 2872 students from 244 high schools (52.3%) took part to the survey. After quality control and consistency checks 2344 observations were retained (1184 boys and 1160 girls) for a total of 153 (32.8%) high schools.

The questionnaire comprises three main parts: a) individual characteristics, b) family background, c) students' school performance and stylized time use. Individual characteristics include sex, year of birth, nationality, religiosity, smoker behavior and future goals, such as tertiary education enrollment and job aspiration after school completion. Family background characteristics cover information on geographical living area, family composition, parental educational, type of occupation and sector of the economic activity and family income. Students' school performance comprises information about the math point average (MPA), grade point average (GPA), and experience of grade retention. Our survey contains also information on how many minutes per day the student spends on studying at home in a typical school day. We also ask information on the main activity students carry out in their free time.[8]

In our data, MPA is the average of all math grades obtained by the student during the current school year or part of it, while GPA is the average over all subjects' grade obtained at the end of the previous school year as reported by the student.[9] Effort is measured in hours per day that students spend in studying. On average boys and girls are statistically different in terms of both schooling outcomes and effort (Table 1). Girls have slightly higher values of both MPA (6.66 versus 6.49) and GPA (6.93 versus 6.71) than their male counterpart.[10] Girls also study more hours than boys. On average, girls allocate 2 hours and 52 minutes per day to home study, against 1 hour and 43 minutes of boys.

Table 1: Schooling Outcomes and Effort by Gender

	Boys		**Girls**		**t-test**
	Mean	Std.Dev.	Mean	Std.Dev.	t
Schooling outcomes					
Math Point Average (MPA)	6.49	1.32	6.66	1.24	-3.19**
Grade Point Average (GPA)	6.71	0.74	6.93	0.77	-6.93***
Effort					
Study time (daily h:mm)	1.72 (1h:43mm)	1.11	2.86 (2h:52mm)	1.11	-24.94***
Observations	1184		1160		

Notes: $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure 2 shows the distribution of study hours by gender. Interestingly, Table 2 shows that for girls a higher math grade is not positively associated with studying more, while data

[8] The items are sport, reading, social activities, cultural activities, listening music, gaming, and watching television.

[9] In our analysis, schooling outcome grades (MPA and GPA) are measures of both oral and written exams and are reported by the student.

[10] In Italy, grades are awarded on a 10-point scale, six is the passing mark, and teachers rarely award marks on either extreme, 1-3 or 9-10. If students fail to earn the passing mark in any subject, they have to participate in remedial lessons during the summer. After remedial summer study, students seat a remedial exam for each non-passed subject to be promoted to the next class. Failure to earn the passing mark in any three subjects at one time may lead to failure to be promoted, without opportunity for remedial.

for boys show a positive monotonic relation between math achievement and effort. The table also shows two remarkable empirical facts. First, hard-working boys, defined as those students studying more than 3 three hours and half per day, outperform by 0.4 points hard-working girls in math. Second, hard-working boys outperform by 0.76 points lazy boys, defined as those students studying less than one hour per day, while there is no empirical evidence of this pattern for girls. Gender differences in motivations, dutifulness or in self-confidence may partly explain this observation. Women are generally less self-assured than men (Niederle and Vesterlund 2007), also at younger ages (Bian *et al.* 2017). In line with these personality traits, girls may study more because they need to be assured in their knowledge. Unlike girls, boys tend to be more confident in their competences. In the literature there is also wide evidence about the negative effect of math anxiety on math grades and acquisition of math skills especially among girls than boys (Van Mier *et al.* 2019).[11]

Figure 2: Frequency of Effort by Gender

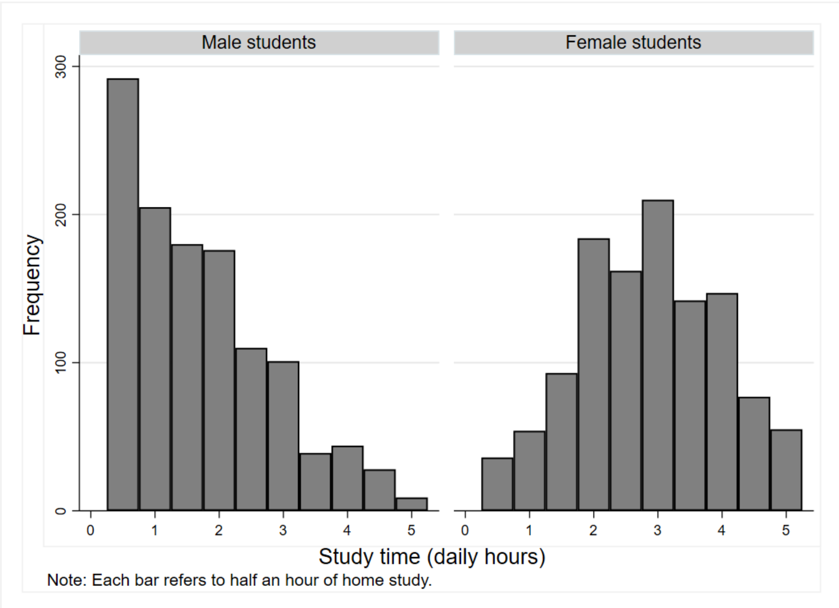


Table 2: Math Point Average (MPA) by Levels of Effort and Gender

	Boys			**Girls**			**t-test**
Effort (study time per day)	Mean MPA	Std.Dev.	%	Mean MPA	Std.Dev.	%	t
$\leq$ 1 hour	6.30	1.32	41.98	6.55	1.34	7.76	-1.61
Between 1h:30mm and 2h	6.45	1.22	30.07	6.70	1.26	23.88	-2.60**
Between 2h:30mm and 3h	6.67	1.38	17.82	6.59	1.24	32.07	0.70
$\geq$ 3h:30mm	7.07	1.27	10.14	6.71	1.19	36.29	2.85**

Notes: **$p < 0.01$.

[11] Descriptive statistics are reported in Table A1 of the Appendix.

4 Empirical Model

We estimate the effect of effort on the production of math achievement (3) for boys and girls, separately, using the following system of two simultaneous equations

$$\begin{cases} Y_{ijp} &= \alpha_1 + \beta_1 X_{ijp} + \delta_1 P_{-ijp} + \psi_1 S_{jp} + \gamma_1 E_{ijp} + \varepsilon_{1ijp}, \\ \\ E_{ijp} &= \alpha_2 + \beta_2 X_{ijp} + \delta_2 P_{-ijp} + \psi_2 S_{jp} + \zeta_2 Z_p + \varepsilon_{2ijp}, \end{cases} \tag{9}$$

where i refers to the student enrolled in school j in province p.

The first equation of the system estimates the mathematical attainment Y of student i, which is measured by the point average of the enrolled school year. As described in Section 2, the production of human capital requires a number of input factors often not collected in dataset. However, our data allow us to address many of the concerns raised in the literature of human capital production. We observe individual and family characteristics X, which are fundamental to explain the production of schooling outcomes, peer attributes P_{-i}, computed by excluding individual i, school characteristics S, and student effort E. ε_1 and ε_2 are two error terms with mean 0, variance σ_1^2 and σ_2^2, and covariance σ_{12}. We use school-level clustering to estimate standard errors to account for correlations across students enrolled in the same school.

The second equation of (9) captures the provision of effort E. The model specification of E includes the same set of individual, family, peer and school characteristics that enters the schooling outcome equation. Even though we control for a number of factors associated with the production of human capital, there can be unobserved heterogeneity correlated with both schooling outcome and effort leading to biased estimation of the coefficient of interest, γ_1. If unobserved student aptitudes are positively correlated with both mathematical attainment and effort, then OLS estimates of the parameter γ_1 are biased upward. On the other hand, if unobservables are negatively correlated with Y and E, then OLS estimates of γ_1 are biased downward. We address the issue of omitted variable bias by introducing a set of instrumental variables Z in the effort equation. They are expected to impact Y indirectly, to the extent that they affect students' effort.

Our set of instruments Z_p includes the 2008 gender-specific unemployment rate of individuals aged 15-24 for the Veneto provinces (Figure A1 in the Appendix) and the average daily rainfall intensity during the school year 2008/2009 (Figure A2 in the Appendix). The unemployment rates, which come from the national statistics (ISTAT 2008), captures the local job market prospects of students to the extent to which a higher unemployment rate is associated with a bad employment outlook (Kalenkosky and Pabilonia 2012, Chadi *et al.* 2019). Current local labour market conditions may not affect students' actual employment status in the future. However, current local labour market conditions may have an impact on students' perception of them (Chadi *et al.* 2019). In turn, this expectation can affect current effort provision. In line with the theory of Section 2, a positive relationship between local unemployment rate and effort provision is expected for a *motivated* student, while a negative relationship is expected for a *discouraged* student. Given that our sample is made by high school leavers and some of them may intend to enrol in university (60% of girls and 51% of boys), instead of entering immediately the labour market, one may argue that students, who intend to pursue further education, might be less affected by the local labour market conditions at the time of school graduation, because they will search for a job later in time. Hence, we perform a robustness analysis to verify whether and how the effect of the local unemployment rate is different for students who intend and do not intend to enrol in tertiary education.

The other instrumental variable is the local intensity of rain precipitation. Empirical evidence about the effect of weather conditions on functional cognition is mixed. Keller *et al.* (2005) find that bad weather has a negative impact on mood and, hence, on executive functions. Differently, Lee *et al.* (2014) assess that bad weather raises worker productivity because individuals are less distracted by those outdoor activities that can be undertaken on good weather days. Similarly, Simonsohn (2010) shows that bad weather reduces the opportunity cost of out-of-home activity participation, such as outdoor sports or hanging out with friends, and may increase the attractiveness of schooling activities. To derive information on weather, we match the residence of students with the nearest weather station.[12] We use the information on millimeters of rain and the number of rainy days of each month between September 2008 and June 2009 to calculate the average daily rainfall intensity during the reference period.[13] We argue that the exposure to a higher rainfall intensity would discourage students from doing outdoor activities, forcing them to stay indoor and likely spend more time in studying activity. We expect a positive effect of the average daily rainfall intensity on effort provision.

The set of individual characteristics X includes religiosity, smoking behaviour, intention to enrol in tertiary education, ambition about future employment (dummy variable equal to one if the student strives for a high-status job), and end-of-fourth-year GPA of the student, which is likely to capture both the influence of historical inputs and innate ability, giving the results a "value-added" interpretation (Todd and Wolpin 2003, Maasoumi *et al.* 2005, Eren and Henderson 2011, Jackson 2018). The idea is that current school achievement is due to an initial endowment of mental capacity (Todd and Wolpin 2003) and the cumulative effect of family, school and peer inputs on students' incoming ability (Jackson 2018).[14] We also control for sport practice and gaming and internet usage as the main activities carried out by the student in her leisure time.[15]

[12] In total, we have the information of 108 weather stations. Information on other climate conditions, such as temperature, are also available. There exists empirical evidence about the negative impact of hot temperatures on cognitive ability and schooling outcomes due to the temperature sensitivity of brain (Graff Zivin *et al.* 2018). Recently, Alberto *et al.* (2021) investigated the leisure-education trade-off as a result of temperature change. However, in Veneto the hottest temperatures are reached in July and August when high schools are closed for summer holidays. In addition, variability in temperatures is low across Veneto provinces, particularly during the spring season.

[13] In Italy the school year begins in mid September and ends in mid June.

[14] Todd and Wolpin (2003) argue that the value-added specification of the EPF is more convincing than that based only on contemporaneous inputs. However, controlling for a lagged measure of school achievement makes the model susceptible to endogeneity bias when information on relevant family and school inputs are missing even if these omitted inputs are orthogonal to the included inputs. This would produce bias estimates of the lagged variable and of all the contemporaneous inputs. The estimation of a causal effect of family and school inputs on schooling outcomes is not the goal of this work, which is instead interested in the causal link between effort provision and schooling achievement.

[15] There is evidence in the literature that participation in sport is positively associated with math and science grades, independently of unobserved individual ability and level of motivation (Lipscomb 2007). A positive association between participation in sport and wages has also been found (Barron *et al.* 2000, Kuhn and Wein-

Further, we control for students' attitudes towards risk by introducing a dummy variable equal to one if the student is *risk lover*, and time preference using a dummy variable equal to one if the student is *forward-looking*. Risk and time preferences have been proved to be systematically related to cognitive abilities (Dohmen *et al.* 2010) and predict a wide range of relevant economic outcomes, as well as choices about investment in education (Dohmen *et al.* 2018). Individual risk aversion and patience are measured through lottery experiments (Coller and Williams 1999, Harrison *et al.* 2002, Holt and Laury 2002).

The socio-economic characteristics of the family are relevant factors of the educational production function. These characteristics includes the geographical location of the household (urban vs rural areas), family structure (divorced parents and number of siblings),[16] family income (dummies for income quintiles), home ownership and parental human capital (mother's and father's tertiary education, the mother is a teacher). The educational attainment also depends on parental effort. Parents may directly help with homework, or provide educational experiences that increase interest for study of the child. Parental effort is captured by parents' involvement in the school life of the child. The variable is a dummy equal to one if the student made the decision whether to pursue or not further education mainly discussing with parents.

Sharing a classroom environment with high-ability classmates might positively affect either the process of gaining skills or the motivation of working harder, or both. As the presence of "bad apples" within classrooms has been proved to have negative effects on other classmates' schooling outcome (Lazear 2001), there is also empirical evidence that the gender composition of classes produces significant effects on schooling outcomes (Hoxby 2000, Whitmore 2005, Lavy and Schlosser 2011, Hill 2015). In general, a higher proportion of females has positive effects on peers' mean test grades, which would work through a reduction in classroom violence and disruption, besides improving the inter-students and student-teacher relationships (Lavy and Scholsser 2011). Our data allow us to define two measures of peer effects P_{-ij}: 1) the average GPA of peers, which is an *endogenous peer effect* (Manski 1993) because individual i is herself a peer of her peers and influences each other, and 2) the proportion of female peers in the classroom, which is an *exogenous peer effect*.[17] We opt for a linear-in-means model, where the peer effect is measured through the mean value of a peers' characteristic excluding the individual.

School quality is another important input in the production of human capital.[18] We adopt

berger 2005). Mixed evidence exists on the nexus between the use of internet/electronic gaming and educational outcomes (Posso 2016).

[16] McLanahan and Sandefur (1994) show that children who grew up in single-parent families or with step-parents have lower educational achievements than those who grew up in two-biological-parent families. Ginther and Pollak (2004) find that children reared in traditional nuclear families have better educational outcomes than children reared in single-parent or blended families. However, when they control for family characteristics, such as income, the correlation between family structure and children's educational outcomes becomes negligible. Kalenkoski and Pabilonia (2012) find that family structure is correlated with the degree of parental supervision, which may affect children's time allocation among different activities.

[17] Exogenous and endogenous peer effects cannot be separately identified (Manski 1993). However, ignoring endogenous peer effects will result in bias estimates of the exogenous peers effect (Moffitt 2001). For this reason, we include both endogenous and exogenous peer effects. Notice that our analysis does not suffer from the so-called *sorting effect* (Sacerdote 2011, Ryan 2017), which is the self-selection into classroom, because in Italy once the school choice is done students are randomly assigned to classroom.

[18] The Italian school system is financed by the central government and, in part, by regional and municipal authorities. Local authorities are responsible for the provision of educational services, such as transport, school meals, and books, only for elementary schools. Teachers are enrolled through national recruitment. Rules

an index of school quality S_{jp} measured by the Fondazione Giovanni Agnelli (FGA) for each Italian high school and provided us by the Agnelli Foundation's Eduscopio project (Aina *et al.* 2019). The FGA index gives a measure of the effectiveness of the high school in preparing students for university education. It is the average of two indicators, equally weighted. The first indicator is the average mark obtained at university exams of the first year. The second is the percentage of university training credits obtained in the first year.[19] Both indicators refer to the results achieved at the first year of tertiary education because a good high school should favour a better transition to university. While, after the first year, success in studies is increasingly affected by the experience gained in the field and less by the origin of the school. The FGA index is highly correlated with the type of high school (lyceum, professional and vocational school), where lyceums have a higher FGA than the other high schools. However, within the same school type, there is a significant difference in the FGA showing that quality is not fully captured by the type of high school.

5 Results

The exposition of the results follows the estimation roadmap. We first present in section 5.1 the OLS estimates. Section 5.2 discusses the IV estimates to control for confounding effects and test for weak instruments. Because tests are not conclusive about the strength of our instruments, in Section 5.3 we form identification-robust confidence sets (Andrews 2018, Andrews *et al.* 2019) for our IV estimates. In Section 5.4 we conduct an instrument-free inference under bounded regressor endogeneity (Kiviet 2020) to verify the validity of the results. In section 5.5 we investigate in more detail the role of family background in effort provision and math achievement.

5.1 OLS Estimation

The econometric analysis starts with the estimation of the production of school achievements using Ordinary Least Squares (OLS). An initial interesting finding is that there are remarkable differences in the factors explaining math performance between boys and girls (Table 3).[20] OLS estimations reveal a non significant association between MPA and study hours for female students, suggesting that studying more does not increase the math performance of girls. Unexpectedly, looking at the adjusted R^2, effort does not predict the math attainment of girls. The adjusted R^2 is null when controlling only for effort (Panel B, column 1 of Table 3), while it increases when other explanatory variables are added into the model specification, reaching its highest value of 0.268 when controlling for the end-of-fourth-year GPA (column 5). Thus, GPA is a strong predictor of the math outcome of girls.

employed to assign teachers to schools are strongly based on seniority (Bryson *et al.* 2020). In our sample, private high schools are 7 out of 153.

[19]In the Italian university grading system training credits correspond to the European Credit Transfer and Accumulation System.

[20]For convenience, we report only the estimates of the key explanatory variable, study hours. For the full set of results see Table A2 in the Appendix.

Differently, the boys' investment on an additional hour of home study is in general associated with a significant increase in MPA grade. For instance, controlling only for effort (Panel A, column 1), one extra hour of studying increases MPA by 0.21 points on a 10-point scale. When controlling for individual and family characteristics, peers' attributes and school quality, the effect associated with study hours diminishes (columns 2 and 3). The significant reduction in the explanatory power of study hours may suggest the relevance of these variables also in explaining how much effort students put into studying, besides capturing the effect of some unobservables related to effort provision. When among other variables end-of-fourth-year GPA is controlled for (column 5), the correlation between MPA and study hours loses statistical significance. In line with the evidence found for girls, the end-of-fourth-year GPA, which can be a proxy for historical schooling inputs, accumulated knowledge over previous grades and endowed ability, plays a major role in explaining individual current math achievements.

The results of Table 3 cannot be interpreted as causal. In our setting, the major peril to causal identification is omitted variable bias. For instance, we do not observe individual numerical, verbal and general reasoning aptitudes, or other factors that can affect MPA and study hours confounding the correlations in our OLS estimates. In the next section we deal with the endogeneity issue and discuss IV estimates of the educational production function.

Table 3: Production Function of Math Achievement by Gender: OLS Estimation

	Math Achievement (MPA)				
	(1)	(2)	(3)	(4)	(5)
Panel A. Boys					
Effort (study hours per day)	0.206***	0.090**	0.083**	0.082**	-0.002
	(0.042)	(0.037)	(0.039)	(0.040)	(0.033)
Observations	1184	1184	1184	1184	1184
Adjusted R^2	0.029	0.115	0.115	0.114	0.337
Panel B. Girls					
Effort (study hours per day)	0.026	-0.008	-0.001	0.019	-0.030
	(0.031)	(0.030)	(0.031)	(0.032)	(0.034)
Observations	1160	1160	1160	1160	1160
Adjusted R^2	-0.000	0.039	0.039	0.047	0.268
Controls					
Individual and family characteristics		X	X	X	X
Peers' characteristics			X	X	X
School quality				X	X
Past school performance: GPA					X

Notes: Please refer to Section 4 for the description of control variables. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 IV Estimation

Table 4 shows the first-stage estimation of students' effort decision. Empirical findings suggest that students' decision about how much effort put into studying is substantially different by gender when individuals confront them with bad job market prospects and obstacles in entering paid work after school completion. A worsening in labour market opportunities, due to a raise in the local unemployment rate, is associated with an increase in study hours for boys and a decrease for girls. Hence, boys tend to behave as *motivated* students, while girls as *discouraged* students.[21] Differences in gender attitudes towards difficulties and failures are not new in the literature. As for education, a number of studies have shown that while male students tend to attribute their failures to lack of effort, girls hold lack of ability responsible for their failures (Nicholls 1975, Reis 1987). In terms of labour market participation, women are generally more likely than men to be discouraged workers (Baussola and Mussida 2014).[22]

Table 4: Estimation Results for the Effort Equation (IV First-stage Estimates)

	Boys		**Girls**	
Instrumental variables				
Unemployment rate	0.032**	(0.014)	-0.031*	(0.016)
Rainfall intensity	0.023**	(0.009)	0.004	(0.012)
Individual and family characteristics				
Religious	0.141	(0.142)	0.182	(0.160)
Free time: Sport	-0.167*	(0.085)	-0.034	(0.093)
Free time: PS and Internet	-0.277***	(0.092)	-0.103	(0.156)
University enrolment	0.239***	(0.088)	0.172*	(0.100)
Career ambition	0.183***	(0.053)	-0.091	(0.056)
Smoker	-0.202***	(0.061)	-0.162**	(0.077)
Patient	0.106	(0.065)	-0.007	(0.065)
Risk lover	0.036	(0.076)	-0.035	(0.069)
Mother: highly educated	0.065	(0.121)	-0.145	(0.140)
Mother: teacher	0.277*	(0.144)	0.033	(0.182)
Father: highly educated	0.111	(0.093)	0.008	(0.090)
Divorced parents	-0.126	(0.132)	-0.267**	(0.130)
Number of siblings	-0.033	(0.036)	-0.094***	(0.028)
Family income: 2nd quintile	-0.139	(0.104)	-0.041	(0.088)
Family income: 3rd quintile	-0.131	(0.122)	-0.127	(0.096)
Family income: 4th quintile	-0.105	(0.105)	-0.212**	(0.103)
Family income: 5th quintile	-0.181	(0.109)	-0.242**	(0.104)
Home owner	-0.037	(0.101)	-0.044	(0.079)
Urban	-0.048	(0.067)	-0.092	(0.071)
Parental effort	0.181***	(0.067)	0.107*	(0.062)
Peers' characteristics				
Share of female classmates	0.525***	(0.169)	0.125	(0.214)
Classmates' average GPA	0.301**	(0.140)	0.416**	(0.189)
School quality				
School's FGA index	0.016***	(0.005)	0.023***	(0.005)
Past school performance				
Own end-of-fourth-year GPA	0.200***	(0.049)	0.138***	(0.042)
Observations	1184		1160	
Adjusted R^2	0.213		0.113	
F statistic for excluded instruments	5.45		2.60	
p-value of F statistic	0.006		0.081	

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Weather conditions have different effects across gender. Male students tend to be more diligent during bad weather days, while girls' effort is not affected by rainfall intensity. The evidence

[21] Given that students who intend to enrol in university might be less concerned by the local labour market conditions, because they will search for a job later in time, we test the robustness of our results by estimating the effort equation on the subsample of high-school students who intend to enrol in tertiary education and on the subsample of those who do not want to pursue further education. On the entire sample of students, we also introduce an interaction between the local unemployment rate with the dummy variable for the intention to enrol in university. Empirical results, showed in Table A3 in the Appendix, suggest that the extent of the effects associated with the local labour market conditions are comparable in the two subsamples and consistent with the results in Table 4, but they are less precisely estimated. In line with these estimates, the interaction term, estimated on the entire sample of students, is not significantly different from zero.

[22] In 2009, in Italy 9.1% of young women aged 15-24 were available to work but not actively seeking, against 8.3% of young men (see Eurostat, Supplementary indicators to unemployment by sex and age [lfsa_sup_age]).

for boys can be due to a reduced opportunity-cost of outdoor activity participation during bad weather days. Generally, boys are more likely than girls to undertake out-of-home activities, such as meeting friends, or practicing outdoor sports.[23] An alternative explanation may be related to girls' higher diligence. Otherwise, girls' study habits are not, or only marginally, influenced by the weather, being girls more likely than boys to be hard-working students.[24]

Boys and girls are different also under other respects. Boys with high career ambitions tend to work harder, but this is not the case for girls. Both male and female students who intend to enrol into tertiary education exert more effort, but the extent of the effect is larger for boys. These results seem to suggest that the nexus between effort provision and a personality trait such as *being ambitious* is stronger for boys than for girls. In other words, boys should have important motivations and ambitions for their future to work harder at school. Differently, girls are committed at school whatever their projects and ambitions for the future. Other individual traits, such as risk and time preferences, are not significantly associated with students' effort provision.

Family characteristics are more important for girls than for boys. Girls who live in a disrupted family or who belong to a wealthy household tend to be less committed to studying. These effects are negative also for boys but not statistically significant. Parental effort matters to a larger extent for boys than for girls.

Peers' characteristics and school quality play an important role for students' effort provision. The higher the mean GPA among classmates, the larger the number of hours devoted to home study for both boys and girls. This can be due to an imitation effect. In other words, having high-ability classmates might positively affect students' motivation to work harder. For boys also the gender composition of high school classmates matters. An increase of one standard deviation in the share of female classmates is associated with an increase of about 8 minutes of study time per day. This evidence may be suggestive that the presence of female classmates favours the development of a more disciplined and motivated behaviour among male students. As for the effect related to school quality, for both boys and girls, an increase of one standard deviation in the FGA index is associated with an increase of about 13 minutes of study time per day for males and 8 minutes for females. Hence, a good school environment is likely to foster students' motivation to work harder.Lastly, both boys and girls, who achieved a higher end-of-fourth-year GPA, are more likely to study more hours. It may be that GPA is a proxy for past choices on effort and a student who exterted more effort in previous grades is also likely to put more effort at the year of graduation.

Table 5 shows IV estimates of math achievement for boys and girls. For boys a one-hour increase in daily study time results in a 1-point increase in MPA on a 10-point scale. The

[23] Data from the Italian National Institute of Statistics (ISTAT) reveal that boys are considerably more likely than girls to meet friends in their freetime. In 2009, 64.1% of boys in the age group 18-19 was used to meet friends every day and 26.5% at least once a week, against 48% and 32.3%, respectively, for girls (http://dati-giovani.istat.it/#). Furthermore, ISTAT's *Indagine "Cittadini e il tempo libero"* shows that 71.9% of boys aged 18-19 practice sports against 47.7% of female peers (ISTAT 2015).

[24] The value of the F statistic for excluded instruments, which tests the jointly significance of the unemployment rate and the rainfall intensity in the effort equation, is equal to 5.45 for boys and 2.60 for girls. The null hypothesis of jointly significance of our instruments in the effort equation is not rejected.

effect is sizeable and larger than OLS estimate suggesting that students with high numeracy aptitude could achieve higher math grades with relatively less effort. In other words, it could exist a negative correlation between effort and unobserved numeracy ability for the subsample of abler students. In line with our evidence, Stinebrickner and Stinebrickner (2008) find that more effort can make up for lower cognitive ability. Similar, Chadi *et al.* (2019) find a negative effect of ability on effort and speak about a "lazy genius phenomenon." For girls, in line with OLS estimates, the effect of effort on MPA is not statistically different from zero. Studying more does not help girls to achieve better scores in math. These results suggest that, for girls, the road to obtain better achievements in math is not encouraging and promoting diligence and committment. It is important to implement programs that make girls aware that math is not a male domain. Because math skills are not innate, improving girls' self-confidence, and passion for the many beauties of math may prove more effective.

Table 5: Production Function of Math Achievement by Gender (IV Second-stage Estimates)

	Boys		**Girls**	
Effort				
Study hours per day	0.979**	(0.430)	-0.260	(0.387)
Individual and family characteristics				
Religious	0.020	(0.210)	0.136	(0.219)
Free time: Sport	0.235*	(0.123)	0.144*	(0.083)
Free time: PS and Internet	0.345**	(0.160)	0.135	(0.191)
University enrolment	-0.050	(0.158)	-0.052	(0.101)
Career ambition	-0.069	(0.094)	0.123*	(0.075)
Smoker	-0.005	(0.123)	-0.123	(0.098)
Patient	-0.038	(0.104)	0.043	(0.072)
Risk lover	0.095	(0.129)	-0.107	(0.077)
Mother: highly educated	-0.123	(0.149)	-0.015	(0.150)
Mother: teacher	-0.369	(0.243)	-0.029	(0.194)
Father: highly educated	0.074	(0.147)	0.213*	(0.111)
Divorced parents	0.149	(0.170)	-0.198	(0.193)
Number of siblings	0.125**	(0.052)	0.003	(0.055)
Family income: 2nd quintile	0.121	(0.148)	0.123	(0.107)
Family income: 3rd quintile	0.157	(0.158)	0.272***	(0.103)
Family income: 4th quintile	0.076	(0.161)	-0.007	(0.129)
Family income: 5th quintile	0.253	(0.182)	0.101	(0.145)
Home owner	-0.016	(0.125)	-0.019	(0.089)
Urban	-0.030	(0.104)	-0.043	(0.071)
Parental effort	-0.269**	(0.121)	-0.042	(0.088)
Peers' characteristics				
Share of female classmates	-0.441	(0.419)	0.023	(0.355)
Classmates' average GPA	-0.354	(0.231)	-0.218	(0.226)
School quality				
School's FGA index	-0.018	(0.011)	-0.001	(0.011)
Past school performance				
Own end-of-fourth-year GPA	0.734***	(0.099)	0.844***	(0.075)
Observations	1184		1160	
Centered R^2	-0.187		0.246	
Diagnostics				
Wu-Hausman F test (test of endogeneity of study hours)	16.52		0.60	
p-value of Wu-Hausman F statistic	0.000		0.432	
Kleibergen-Paap rk LM statistic (underidentification test)	19.53		9.77	
p-value of Kleibergen-Paap rk LM statistic	0.000		0.008	
Hansen J statistic (overidentification test of all instruments)	0.029		6.271	
p-value of Hansen J statistic	0.865		0.012	
Kleibergen-Paap rk Wald F statistic for weak identification	9.99		4.92	

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For both boys and girls the other control variables have almost no explanatory power, except for own end-of-fourth year GPA, which is strongly and positively correlated with math achievement. This suggests that family background and structure, peers' characteristics and school quality are more likely to be related with students' behaviour in terms of effort provision (Table 4), rather than having a direct association with math outcome (Table 5). However, this does not mean that they are not important for individual's schooling attainment. They are relevant in that they are related with students' decision about how much effort put into studying, which in turn, affects math grade, at least for boys. For girls the story is different.

The Wu-Hausman F test indicates that study hours are endogenous in boys' MPA equation, while they are exogenous in girls' outcome equation. The Kleibergen-Paap rk LM statistic suggests that the (excluded) instruments are correlated with the endogenous regressor, for both boys and girls.[25] The Hansen J test of overidentifying restrictions[26] is not rejected for boys, while it is rejected for girls. For female students the rejection casts doubt on the validity of our instruments. In addition the value for the Kleibergen-Paap *rk* Wald F statistic for weak identification is quite low.[27] However, the effect of study time on MPA is not significant. The evidence for boys is more comforting. The value of the first-stage F-statistic is 9.99, which is very close to the Staiger and Stock (1997) rule-of-thumb cutoff of 10. Given that our instruments are not weak, but not strong either, to resolve the resulting ambiguity, in the next subsection we construct robust confidence sets for our IV estimates.

To account for the ordinal nature of math grades, which range from 0 to 10, we estimate an IV ordered probit model by full information maximum likelihood. The outcome variable Y in system (9) is now defined as an ordinal categorical variable which takes value 1 if the student's MPA is below 6, which is the passing mark, 2 if the MPA ranges from 6 to 7.5, 3 if MPA is equal or larger than 8. Table 6 shows the main empirical findings.[28] The effects of our instrumental variables on study hours are similar to the ones presented in Table 4. In line with our previous estimates, studying more does not help girls to achieve better grades in math. On the other hand, for boys an additional daily hour of home study is important to have an above average MPA. Remarkably, it raises the probability of obtaining a very good grade in math (MPA larger than 8) by about 20 percentage points, while reducing by roughly the same amount the probability of not obtaining the passing mark. These results are further evidence that effort is a relevant input only in boys' educational production function.

[25]The LM statistic tests whether the equation is identified, that is whether the excluded instruments are relevant, meaning correlated with the endogenous regressors.

[26]The null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.

[27]For girls, we also run the IV estimates using only the unemployment rate as (excluded) instrument. The estimated coefficient (standard error) associated with the unemployment rate is -0.032 (0.015), significant at 5% level. The Wu-Hausman F test indicates that effort is exogenous in MPA equation and the Kleibergen-Paap *rk* LM statistic suggests that the unemployment rate is correlated with study hours. The Kleibergen-Paap *rk* Wald F statistic for weak identication is 11.41, relatively close to 10, the common rule of thumb cutoff for weak instruments. Study hours continue to be not significant in MPA equation.

[28]The full set of results are available from the authors upon request.

Table 6: Marginal Effects of an Instrumental Variable Ordered Probit Model

	Boys				Girls			
Dependent variable	Effort	Category of MPA			Effort	Category of MPA		
		1	2	3		1	2	3
Panel A. First-stage estimates (Effort equation)								
Instrumental variables								
Unemployment rate	0.028**				-0.032**			
	(0.014)				(0.015)			
Rainfall intensity	0.026**				-0.000			
	(0.009)				(0.014)			
Panel B. Second-stage estimates (MPA equation)								
Outcome probability		0.346***	0.388***	0.266***		0.254***	0.501***	0.245***
		(0.049)	(0.095)	(0.050)		(0.053)	(0.120)	(0.068)
Effort								
Study hours		-0.203***	0.038***	0.165***		0.106	-0.003	-0.103
		(0.053)	(0.007)	(0.040)		(0.118)	(0.005)	(0.120)
Observations	1184	324	633	227	1160	263	647	250

Notes: Categories 1, 2 and 3 refer to MPA in the range [4; 5.5], [6; 7.5] and [8; 10], respectively. Please refer to Section 4 for the description of control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3 Weak-instrument Robust Confidence Sets

Weak identification arises when the correlation between the instrumental variable and the endogenous regressor is small. When weak identification is suspected, researchers should use weak-instrument robust methods for inference and form identification-robust confidence sets, which can be informative even in presence of weak instruments. These robust confidence intervals have the exact size α independent of the strength of the instrumental variables. They are constructed by inverting robust tests to weak instruments.

In over-identified models, Conditional t, Kleibergen score, and Conditional Likelihood Ratio (CLR) tests are efficient under strong instruments (Andrews *et al.* 2006). However, the authors recommed the use of the CLR test because it is nearly optimal. While there is consensus in the literature that CLR has very good power properties for homoskedastic settings (with single endogenous regressor), this is not straightforward in non-homoskedastic ones. In the latter case the literature has not yet converged on what procedures to use in practice. Andrews *et al.* (2019) recommend researchers to use procedures that are efficient when the instruments are strong. Following their suggestions, we consider the CLR test which is based on the conditional approach proposed by Moreira (2003) and uses LR statistic. The CLR confidence set (CLR-CS) may be a finite interval, a union of two infinite intervals or the whole line. The latter appears mainly in presence of weak instruments. In this case, there is little or no information about the coefficient of interest and this is correctly shown by the structure of the confidence set (Mikusheva 2010).

To complement the analysis, we also form the two-step identification-robust confidence set (LC 2SLS-CS) with controlled coverage distortion (Andrews 2018). A linear combination test is used to form the LC 2SLS-CS. In computing the robust confidence interval we set $\alpha = 5\%$ and a desired level of coverage distortion $\gamma = 10\%$, that parametrises our tolerance for weak instruments.[29] The choice of γ affects the size of the LC 2SLS-CS. When instruments are strong, the robust LC 2SLS-CS with coverage level $1 - \alpha - \gamma$ should be contained in the non-robust Wald confidence set (Wald-CS) with coverage level of $1 - \alpha$, that is reliable when the model is

[29]The reasearcher can chose a value of γ according to her preference for coverage distortion.

well identified. The strength of identification can be assessed by checking how larger γ needs to be for this containment to hold. Then, the reasearchers should bound coverage distortion accordingly.

Table 7 shows CLR test and CLR-CS,[30] besides the robust LC 2SLS-CS, the non-robust Wald-CS and data-driven distortion cutoff that indicates identification strenght.[31] The tests and weak-instrument robust confidence intervals refer to the IV estimates shown in Table 5. For boys the CLR test is 6.48, and the p-value is 0.017. Hence, we reject the null that effort is not statistically significant in MPA equation at 5% level. This is not the case for girls. For boys, the CLR-CS is a finite interval [0.242; 3.576], while it covers the entire grid for girls indicating that the parameter associated with study hours in girls' MPA equation is poorly identified or unidentified.

Table 7: Weak-IV Robust Inference and Identification-robust Confidence Sets

	Boys	**Girls**
Panel A: weak-IV robust inference and identification-robust confidence sets		
CLR statistic (p-value)	6.84 (0.017)	3.84 (0.095)
CLR-CS	[0.242; 3.576]	[entire grid]
Panel B: Two-step identication-robust confidence sets$^{(a)}$		
LC 2SLS-CS	[0.364; 2.788]	[entire grid]
Distortion cutoff level	22%	21%
Panel C: Wald-CS	[0.139; 1.820]	[-0.985; 0.464]

Notes: (a) The desired level of coverage distortion is set to 10%.

As for the LC 2SLS-CS, the estimated (data-driven) distortion cutoff level is 21% for boys and 21% for girls, which are larger than the distortion level of 10% that we chose as maximum tolerance. These values of the estimated distortion cut-off do not rule out the the possibility of weak instruments. Accordingly, we focus on LC 2SLS-CS, instead of the Wald-CS. For boys, the LC 2SLS-CS is a positive finite interval and the extent is quite comparable to the CLR-CS. For girls, the LC 2SLS-CS covers the whole line, as we have already observed for the CLR-CS.[32]

To summarise, our empirical findings suggest that effort is an important input of the education production function, but only for boys. For the latter, diagnostics do not unambiguously indicate that our instruments are strong. However, the evidence of a closed form of both the CLR-CS and LC 2SLS-CS, which are quite comparable in extent, is comforting. As a sensitivity analysis, in the next section we discuss the results of an instrument-free inference under confined regressor endogeneity (Kiviet 2020).

5.4 Kinky-Least-Squares Inference

Although weak-instrument robust techniques for inference can be informative in presence of weak instruments, they do yield confidence intervals which are quite wide or even unbounded. Recently, Kiviet (2020) has proposed a method to achieve set identification based on credible

[30]We use the Stata Package *weakiv* developed by Finlay et al. (2013).

[31]We use the Stata Package *twostepweakiv* developed by Sun (2018) and based on Andrews (2018). The LC 2SLS-CS does not assume that the data are homoskedastic.

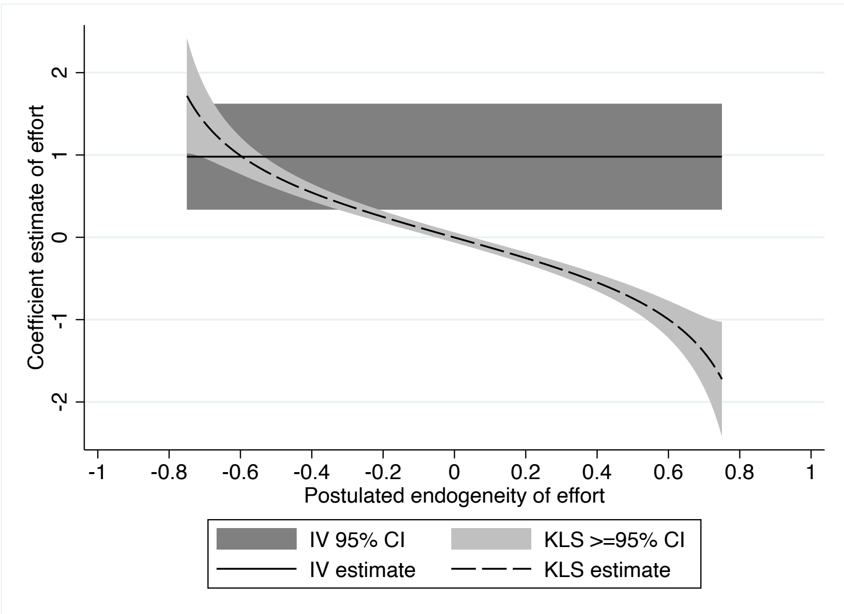
[32]If we are willing to tolerate a coverage distortion of up to 10%, which is less than the distortion cutoff, then we should focus on the LC 2SLS-CS. Differently, if we are willing to tolerate a coverage distortion of up to 30%, which is larger than the distortion cutoff, then we should focus on the Wald-CS.

ranges for the correlation between the endogenous regressors and model errors. Instead of assuming strict orthogonality of instrumental variables and disturbances, the author assumes bounds on the possible non-orthogonality of the endogenous regressors and disturbances. In other words, the author makes some assumptions about the degree of regressor endogeneity and places bounds directly on the endogeneity correlation itself, $\rho_{E\varepsilon_1}$. Even though the latter is unknown, inference can be obtained over a range of some chosen realistic values $r_{E\varepsilon_1}$.[33] If actual endogeneity respects the specified bounds, then asymptotically valid instrument-free inference on coefficients will be obtained.

In deriving the instrument-free least-square estimator, which is addressed as kinky-least-squares (KLS), Kiviet (2020) assumes that sample observations are independently and identically distributed and variables are either normally distributed or have at least no excess kurtosis. Although based on these assumptions, the approach can be very useful for a sensitivity analysis, by providing extra statistical evidence.[34]

We run this sensitivity analysis only for boys, given that previously discussed results for girls suggest that effort does not significantly affect math outcomes and we have a clearcut problem of weak instruments and identification, even though the Wu-Hausman F test does not reject the null that effort is exogenous in the girls' outcome equation. In Figure 3, KLS point estimates[35] for the effect of study time on boys' math grade and related confidence set are plotted over a wide range of endogeneity correlations between study time and model disturbances.

Figure 3: KLS and IV Point Estimates and Confidence Sets for Effort of Boys



[33]The value of $\rho_{E\varepsilon_1}$ lies in the interval $\left(r^L_{E\varepsilon_1}, r^U_{E\varepsilon_1}\right)$ with $-1 < r^L_{E\varepsilon_1} < \rho_{E\varepsilon_1} < r^U_{E\varepsilon_1} < 1$.

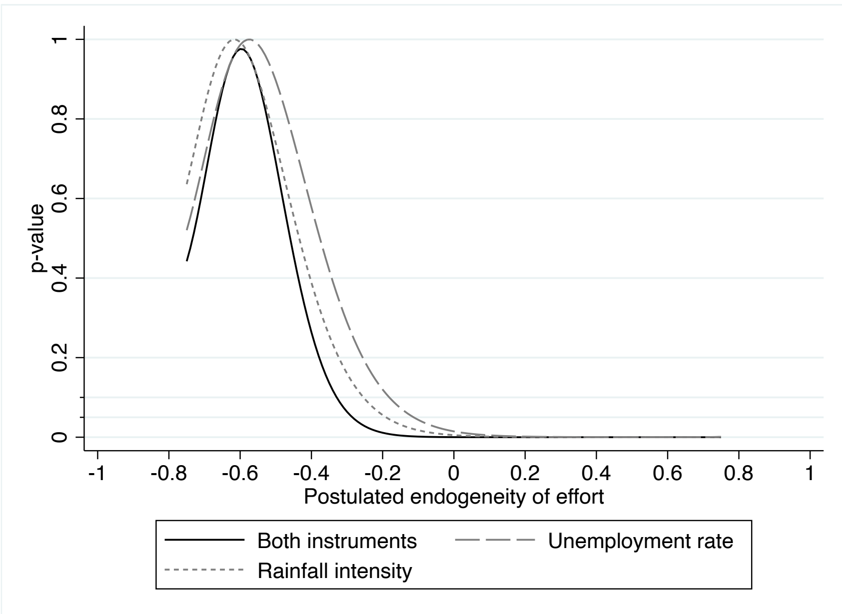
[34]Other authors developed alternative instrument-free identication strategies by imposing bounds on some aspects of the relationship between observable and unobservable model characteristics. For instance, Altonji et al. (2005) achieve set identification by imposing bounds on the correlation between the error terms in the outcome and the selection equation. Oster (2019) achieves interval identification by imposing bounds on the relative degree of selection on observables and unobservables and on the coefficient of determination that could be hypothetically obtained if all unobservables were controlled for in the regression analysis.

[35]We use the Stata Package *kinkyreg* developed by Kripfganz and Kiviet (2021).

The estimated correlation between the IV residuals and the effort variable is -0.57. As Kiviet and Kripfganz (2020), we consider the range $(-0.75; 0.75)$.[36] The results are compared to the IV estimate and 95% confidence interval. The KLS confidence set varies with the value of the endogeneity correlation and is narrower than the confidence interval from IV. The latter is instead invariant regarding the value of $r_{E\varepsilon_1}$ and centered at the IV estimate 0.979. OLS yields a coefficient estimate for effort of -0.002, which is obtained supposing $r_{E\varepsilon_1} = 0$. Notice that the KLS findings are in line with IV results for negative values of the endogeneity correlation. The IV confidence interval, which is contingent on the validity of our instruments, conforms in width to the KLS confidence set for $-0.71 < r_{E\varepsilon_1} < -0.54$. IV and KLS results are in sharp contrast for positive values of the endogeneity correlation. Notice that the estimated correlation between the IV residuals and study time is -0.57, which is consistent with the observed range for the endogenous correlation where IV estimate is within the KLS confidence set. This evidence is comforting because in line with the downward-biased OLS estimates. Hence, students with higher numeracy ability or mathematical skills, which are unobservables in our model, can grade very high on math with relatively low levels of effort.

The KLS method allows us to test the validity of instruments by testing exclusion restrictions. Figure 4 plots the p-values of the F tests for the (individual and joint) significance of the instruments' coefficients over the endogeneity range.

Figure 4: P-values of Exclusion Restriction Tests on Instrumental Variables



The maximum p-value for the joint exclusion restrictions test is observed for a value of $r_{E\varepsilon_1}$ in the neighborhood of -0.60, which is very close to our estimates of the endogeneity correlation. The validity of our external instrument lacks strong support for $r_{E\varepsilon_1} > 0$. This means that if we have *a priori* reasons to believe that the correlation between effort and the model errors is negative, then results in Figure 4 provide support for the validity of our instruments. However,

[36] As in Kiviet and Kripfganz (2020)'s empirical applications, the chosen range is within the lower and upper bounds of the endogeneity correlation. In our case, KLS is defined for $|r_{E\varepsilon_1}| < 0.88$.

Kiviet argues that another possible interpretation of the graph is possible. If one has *a priori* reasons to believe that the instruments must be valid, then the graph provides information that the true value of the endogeneity correlation should be negative. As a further check, we computed the t-statistic version of the Durbin-Wu-Hausman test. Its negative sign indicates a negative endogeneity correlation.

5.5 The Role of Family Background

We also intend to learn more about the role of family background in effort provision and math outcome. We estimate system (9) after stratifying the sample by their Socio-Economic Status (SES) defined in terms of family income. A family has a *Rich SES* when income is above the median, a *Poor SES* otherwise. Table 8 shows that the effect of the unemployment rate on the decision about how much effort boys put into studying is similar whatever the family background.

Table 8: IV Estimates of the Production Function of Math Achievement by Socio-Economic Status (SES)

	Boys				**Girls**			
	Poor SES		*Rich SES*		*Poor SES*		*Rich SES*	
	Effort	MPA	Effort	MPA	Effort	MPA	Effort	MPA
Instrumental variables								
Unemployment rate	0.036*		0.033**		-0.045**		-0.018	
	(0.021)		(0.014)		(0.021)		(0.019)	
Rainfall intensity	0.006		0.034***		-0.014		0.019	
	(0.013)		(0.011)		(0.018)		(0.013)	
Effort								
Study hours		0.471		1.051***		-0.351		0.214
		(0.576)		(0.371)		(0.322)		(0.524)
Individual and family characteristics								
University enrolment	0.370***	0.047	0.122	0.042	0.064	-0.095	0.308**	-0.195
	(0.117)	(0.266)	(0.110)	(0.170)	(0.132)	(0.106)	(0.128)	(0.205)
Career ambition	0.265***	-0.040	0.130**	-0.011	-0.101	0.128	-0.094	0.149
	(0.090)	(0.147)	(0.065)	(0.120)	(0.093)	(0.088)	(0.081)	(0.119)
Parental effort	0.273**	-0.192	0.074	-0.150	0.186**	0.015	0.012	-0.039
	(0.119)	(0.176)	(0.080)	(0.129)	(0.090)	(0.117)	(0.081)	(0.105)
Peers' characteristics								
Share of female classmates	0.175	0.036	0.806***	-0.801	-0.158	-0.143	0.499*	0.066
	(0.268)	(0.431)	(0.184)	(0.491)	(0.304)	(0.364)	(0.285)	(0.510)
Classmates' average GPA	0.299	-0.185	0.355*	-0.481	0.596**	-0.315	0.253	-0.154
	(0.271)	(0.278)	(0.189)	(0.325)	(0.279)	(0.268)	(0.197)	(0.257)
School quality								
School's FGA index	0.017*	-0.015	0.014**	-0.014	0.024***	0.005	0.023***	-0.015
	(0.009)	(0.015)	(0.006)	(0.010)	(0.006)	(0.011)	(0.006)	(0.014)
Observations	547	547	637	637	625	625	535	535
Adjusted R^2	0.173	0.189	0.241	-0.268	0.113	0.132	0.126	0.262

Notes: The column labelled Effort reports first-stage estimates (Effort equation). The column labelled MPA report second-stage estimates (MPA equation). Please refer to Section 4 for the description of the full set of control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

For girls, the negative effect of the unemployment rate on effort provision is significant only for those with a *Poor SES*. This result suggests that the discouraged-type behaviour of girls confronted with bad employment outlooks is related to the poor economic and financial situation of their family of origin. In other words, girls coming from a poor family background may perceive to have less opportunities compared to wealthier peers and less chances to succeed when job market prospects are not favourable. This double burden may explain why they reduce their effort provision as the unemployment rate rises. The effect of the rainfall intensity in the boys' effort equation is significant only for those living in a wealthy family. This is reasonable if boys from a rich background are more likely to undertake outdoor costly activities.

Male students from a *Poor SES* and with important career ambitions are more motivated and committed than their peers from a *Rich SE*S. This may be explained both by their desire to climb the social ladder and by their higher consciousness about the importance of effort to reach a life goal. Parental effort is important for both sons and daughters of families with lower socio-economic status, while it does not significantly affect the effort provision of students from wealthier families. Children from a lower background are probably more motivated to repay their parents' financial effort to support their studies. The effect of peers is highly heterogeneous across gender and family type. As expected, given that school in our sample are mostly public and students are randomly assigned to classes, the effect associated with school quality has the same effect independently of the family background of the students.

6 Conclusions

This study estimates the causal effect of study effort on math achievement of boys and girls attending high school. We use a primary dataset collected by the authors comprising information about family background, including family structure, family income, education, the working status of parents, and geographic indicators of residence. The data set also includes relevant measures of students' non cognitive abilities and personal characteristics such as rate of impatience and risk propensity. Our information set records study time, leisure time and schooling outcomes, such as math point average and grade point average. Information about school quality is derived from external data. Inspection of raw data reveals significant gender differences in the use of time, with girls being more hard-working than boys. But, on average, they perform slightly, but significantly, better than boys in terms of both MPA and GPA. Interestingly, when selecting the most diligent students, male students obtain higher grades than female students.

We estimate a structural equation model that combines information on math performance with hours of home study. Our estimation strategy infers causal relations by relying on an instrumental variable approach validated using weak-instruments-robust confidence sets (Andrews 2018, Andrews *et al.* 2019) and novel partial identification techniques (Kiviet 2020). The econometric estimates are also consistent with an IV ordered probit specification.

Our results suggest that the amount of study time positively affects math outcomes, for boys only. Our estimates show that an additional hour devoted to study raise*s* boys' math grade by 1 point on a 10-point scale. Girls are more diligent than boys, but an additional hour of out-of-school study does not improve their math scores. Boys seem more efficient in the use of time. This may be because boys are more involved in sport activities, which often a factor in help to improve concentration. On the other hand, girls may be more math anxious or less confident in their abilities. The verification of these conjectures is left for future research.

Our findings also reveal that peers' characteristics and school quality play a more important role than family background variables and they affect educational attainments more through effort than directly. Further, the results suggest that male and female students behave differently when facing unfavorable job market prospects. Girls tend to be discouraged and reduce effort provision. Differently, male students are more motivated and put more effort in studying. As compared to girls, the supply of effort of boys is relatively more affected by weather conditions.

From a policy perspective, a better understanding of students' effort decisions and their implications for educational outcomes provide useful evidence to inform policy makers responsible for public education. What we learned about the achievements in mathematics of both male and female high-school leavers of the Veneto region, can be used to design programs tailored to high-school students, aiming at reducing gender disparities in the production of mathematical skills and competences and guaranteeing equal opportunities of participation and success in many careers, especially those in the STEM fields. Job carriers are increasingly becoming more concentrated in the STEM sectors that are leading the post-pandemic recovery. There are significant labour market advantages for individuals employed in STEM fields, including relatively higher wages, lower unemployment rates and growing job opportunities. However, women's involvement in these fields is much lower than men. In our data, only 8% of female students who intend to enrol in tertiary education choses a STEM major, as compared to 34% of boys. The gender gap is very wide also among the top-performing students with a math grade equal or greater than 8, which represent almost the 20% of both boys and girls. Figures reveal that 42% of top-performing boys intend to enrol in a STEM major against only 15% of girls. Our results suggest that to improve access of girls to STEM fields in college, it is likely more successful to implement programs that make girls less susceptible to negative stereotypes about women abilities in math, which discourage girls' pursuit of many outstanding careers in the STEM fields. Program should also increase the interest and passion of girls for science and math, by reinforcing their self-confidence and nourishing positive expectations about STEM careers. At the same time, it is crucial to make STEM majors more welcoming for women and future STEM workforce more apt to host also women with generative aspirations.

Acknowledgements

We gratefully acknowledge financial support from the STAREBEI grant. Special thanks go to Maria Luisa Ferreira, Hubert Strauss, Daniele Checchi, Vito Peragine, Veronica Polin, Gabriella Berloffa, Giuseppe Porro, Tommaso Agasisti and participants to the 62nd Annual Conference of the Italian Economic Association (26-29 October 2021) for their insightful comments and suggestions. Special thanks go to Federica Barzi, Elena Dalla Chiara and Nicola Tommasi for data quality controls, to Mario Pomini, Nadia Dallago, Renato Cavedon, Dolores Rizzotto, Silvia Saoncella, Andrea Ongaro and Enrico Prospero for their help in data collection, and to Martino Bernardi and the Agnelli Foundation for providing the data on the FGA index. We also thank all high school students who undertook the survey, the officers of the Veneto region administration (Ufficio Scolastico Provinciale di Verona and Ufficio Scolastico Regionale di Venezia), and the Headmasters of the schools that administered the survey. Remaining errors are sole responsibility of the authors.

Declaration of Competing Interest

No.

References

[1] Aina, C., Bratti, M., and Lippo, E. (2019) The contribution of high schools to university students' academic performance: The case of Eduscopio. Fondazione Agnelli n. 59 (04/2019).

[2] Alberto, I.C., Jiao, Y., and Zhang, X. (2021) Too hot or too cold to study? The effect of temperature on student time allocation. *Economics of Education Review* (Forthcoming).

[3] Altonji, J. G., Elder, T. E., and Taber, C. R. (2005) Selection on observed and unobserved variables: Assessing the eectiveness of Catholic schools. *Journal of Political Economy,* 113(1): 151–184.

[4] Andrews, I. (2018) Valid two-step identification-robust confidence sets for GMM. *The Review of Economics and Statistics*, 100(2): 337–348.

[5] Andrews, I., Stock, J. H., and Sun, L. (2019) Weak instruments in IV regression: Theory and practice. *Annual Review of Economics*, 11: 727–753.

[6] Andrews, D. W. K., Moreira, M. J, and Stock, J. H. (2006) Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica*, 74(3): 715–752.

[7] Barron, J. M., Ewing, B. T., and Waddell, G. R. (2000) The effects of high school athletic participation on education and labor market outcomes. *Review of Economics and Statistics*, 82:409-421.

[8] Baussola, M., and Mussida, C. (2014) Transitions in the labour market: Discouragement effect and individual characteristics. *Labour*, 28(2): 209–232.

[9] Bian, L., Leslie, S. J., and Cimpian, A. (2017) Gender stereotypes about intellectual ability emerge early and influence children's interest. *Science,* 355: 389–391.

[10] Bryson, A., Corsini, L., and Martelli, I. (2020) Teacher allocation and school performance in Italy. IZA DP No. 13669.

[11] Card, D., and Payne, A. A. (2021). High school choices and the gender gap in STEM. *Economic Inquiry*, 59(1): 9–28.

[12] Carneiro, P. and Heckman, J. (2003) Human Capital Policy. *NBER Working Papers* 9495.

[13] Chadi, A., de Pinto, M., and Schultze, G. (2019) Young, gifted and lazy? The role of ability and labor market prospects in student effort decisions. *Economics of Education Review,* 72:66–79.

[14] Coller, M., and Williams, M. B. (1999) Eliciting individual discount rates. *Experimental Economics*, 2(2): 107–127.

[15] Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006) Interpreting the evidence on life cycle skill formation. In Hanushek, E., and Welch, F. (Eds) *Handbook of the Economics of Education*, 1: 697–812.

[16] De Fraja, G., Oliveira, T., and Zanchi, L. (2010) Must try harder: Evaluating the role of effort in educational attainment. *Review of Economics and Statistics*, 92: 577–597.

[17] Di Tommaso, M.L., Maccagnan, A., and Mendolia, S. (2021) The gender gap in attitude and test scores: a new construct of the mathematical capability. *Feminist Economics*, 27(3): 161–187.

[18] Delaney, J.M., and Devereux, P.J. (2019) Understanding gender differences in STEM: Evidence from college applications. *Economics of Education Review*, 72: 219–238.

[19] Delaney, J.M., and Devereux, P.J. (2021) High school rank in Math and English and the gender gap in STEM. *Labour Economics*, 69 (Forthcoming).

[20] Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2018) On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives*, 32(2): 115–134.

[21] Eren, O., and Henderson, D. J. (2011) Are we wasting our children's time by giving them more homework?. *Economics of Education Review*, 30(5): 950–961.

[22] Finlay, K., Magnusson, L., and Schaffer, M. (2013) WEAKIV: Stata module to perform weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models. http://ideas.repec.org/c/boc/bocode/s457684.html

[23] Ginther, D. K., and Pollak, R. (2004) Family structure and children's educational outcomes: Blended families, stylized facts, and descriptive regressions. *Demography*, 41: 671–696.

[24] Graff Zivin, J., Hsiang, S. M., and Neidell, M. (2018) Temperature and human capital in the short and long run. *Journal of the Association of Environmental and Resource Economists*, 5(1): 77–105.

[25] Harrison, G. W., Lau, M. I., and Williams, M. B. (2002) Estimating individual discount rates in Denmark: A field experiment. *American Economic Review*, 92(5): 1606–1617.

[26] Hill, A. (2015) The girl next door: The effect of opposite gender friends on high school achievement. *American Economic Journal: Applied Economics*, 7(3): 147–177.

[27] Holt, C. A., and Laury, S. K. (2002) Risk aversion and incentive effects. *American Economic Review*, 92(5): 1644–1655.

[28] Hoxby, C. M. (2000) Peer effects in the classroom: Learning from gender and race variation. *NBER Working Papers* 7867.

[29] INVALSI (2009) Rapporto Nazionale PISA 2009.

[30] Jackson, C. K. (2018) What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, 126(5): 2072–2107.

[31] Jansen, M., Lüdtke, O., and Schroeders, U. (2016) Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology*, 46: 116–127.

[32] Jiang, X. (2021) Women in STEM: Ability, preference, and value. *Labour Economics* (Forthcoming).

[33] Kalenkoski, C. M., and Pabilonia, S. W. (2012) Time to work or time to play: The effect of student employment on homework, sleep, and screen time. *Labour Economics,* 19: 211–221.

[34] Keller, M., Fredrickson, B., Ybarra, O., Côté, S., Johnson, K., Mikels, J., and Wager, T. (2005) A warm heart and a clear head the contingent effects of weather on mood and cognition. *Psychological Science*, 16: 724 –731 .

[35] Kiviet, J. F. (2020) Testing the impossible: Identifying exclusion restrictions. *Journal of Econometrics,* 218: 294–316.

[36] Kiviet, J. F., and Kripfganz., S. (2020) Reassessment of classic case studies in labor economics with new instrument-free methods. Working paper, mimeo.

[37] Kripfganz, S., and Kiviet, J. F. (2021) kinkyreg: Instrument-free inference for linear regression models with endogenous regressors. *The Stata Journal* (Forthcoming).

[38] Kuehn, Z., and Landeras, P. (2014) The effect of family background on student effort . *BE J. Econ. Anal. Policy*, 14: 1337–1403.

[39] Kuhn, P., and Weinberger, C. (2005) Leadership skills and wages. *Journal of Labor Economics*, 23: 395–436.

[40] Lavy, V., and Schlosser, A. (2011) Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3: 1–33.

[41] Lazear, E. P. (2001) Educational production. *Q. J. Econ.* 116: 777–803.

[42] Lee, J. J., Gino, F., and Staats, B. R. (2014) Rainmakers: Why bad weather means good productivity. *Journal of Applied Psychology*, 99(3): 504–513.

[43] Levitt, S. D., and List, J. A. (2007) What do laboratory experiments measuring social preferences tell us about the real world. *Journal of Economic Perspectives*, 21(2): 153–174.

[44] Levitt, S. D., and List, J. A. (2009) Field experiments in economics: The past, the present, and the future. *European Economic Review*, 53(1): 1–18.

[45] Lipscomb, S. (2007) Secondary school extracurricular involvement and academic achievement: a fixed effects approach. *Economics of Education Review,* 26: 463–472.

[46] List, J. A. (2009) An introduction to field experiments in Economics. *Journal of Economic Behavior & Organization*, 70(3): 439–442.

[47] Maasoumi, E., Millimet, D., and Rangaprasad, V. (2007) Class size and educational policy: Who benefits from smaller classes?. *Econometric Reviews*, 24(4): 333–368.

[48] Manski, C. F. (1993) Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60: 531–542.

[49] McLanahan, S. S., and Sandefur, G. (1994) *Growing up with a single parent: What hurts, what helps.* Cambridge, MA: Harvard University Press.

[50] McMillan, J. H., Myran, S., and Workman, D. (2002) Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4): 203–213.

[51] Mehta, N., Stinebrickner, R., and Stinebrickner, T. (2018) Time-use and academic peer effects in college. *Economic Inquiry*, 57(1): 162–171.

[52] Moffitt, R. (2001) Policy interventions, low-level equilibria, and social interactions. In Durlauf S., and Young, P. (Eds) *Social dynamics*, MIT Press, Cambridge, Massachusetts.

[53] Nicholls, J. G. (1975) Causal attributions and other achievement-related cognitions: Effects of task outcome, attainment value, and sex. *Journal of Personality and Social Psychology*, 31(3): 379–389.

[54] Niederle, M., and Vesterlund, L. (2017) Do women shy away from competition? Do men compete too much?. *The Quarterly Journal of Economics*, 122(3): 1067–1101.

[55] Oster, E. (2019) Unobservable selection and coecient stability: Theory and evidence. *Journal of Business & Economic Statistics,* 37(2): 187–204.

[56] Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment*, 2(1): 69–88.

[57] Posso, A. (2016) Internet usage and educational outcomes among 15-year-old Australian students. *International Journal of Communication,* 10: 3851–3876.

[58] Reis, S. M. (1987) We can't change what we don't recognize: Understanding the special needs of gifted females. *Gifted Child Quarterly*, 31(2): 83–89.

[59] Ryan, C. (2017) Measurement of Peer Effects. *Australian Economic Review*, 50: 121–129.

[60] Roemer, J. E., and Trannoy, A. (2016) Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, 54(4): 1288–1332.

[61] Sacerdote, B. (2011) Peer effects in education: How might they work, how big are they and how much do we know thus far? In Hanushek, E., Machin, S., and Weossman, L. (Eds) *Handbook of the Economics of Education*, Volume 3, Elsevier, Amsterdam.

[62] Simonsohn, U. (2010) Weather to go to college. *The Economic Journal*, 120: 270 –280.

[63] Speer, J.D. (2017) The gender gap in college major: Revisiting the role of pre-college factors. *Labour Economics,* 44: 69–88 .

[64] Staiger, D., and Stock, J. H. (1997) Instrumental variables regression with weak instruments. *Econometrica*, 65(3): 557–586.

[65] Stinebrickner, R. and Stinebrickner, T. (2008) The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy*, 8(1): Article 14.

[66] Sun, L. (2018) Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models. *The Stata Journal*, 18(4): 803–825.

[67] Todd, P.E., and Wolpin, K.I. (2003) On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113: F3–F33

[68] Van Mier, H. I., Schleepen, T. M. J, and Van den Berg, F. C. G. (2019) Gender differences regarding the impact of math anxiety on arithmetic performance in second and fourth graders. Frontier in Psychology, 9, 2690.

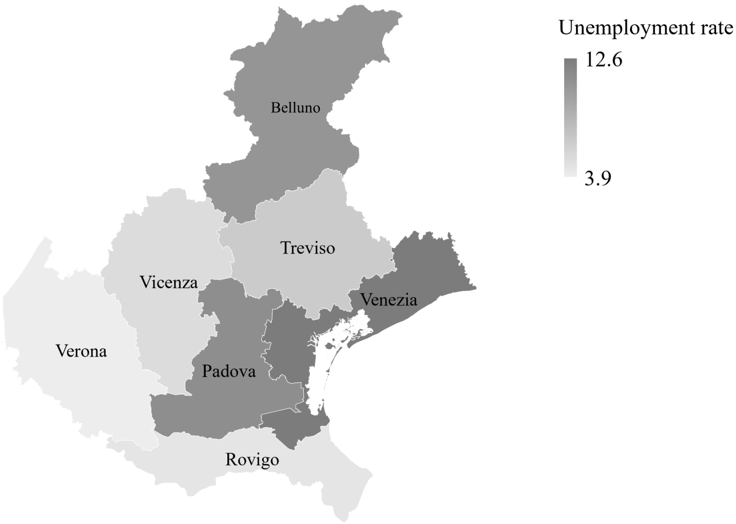
[69] Whitmore, D. (2005) Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *Am. Econ. Rev. Papers and Proceedings,* 95: 199–203.

[70] Wößmann, L., Propper, C., and Duflo, E. (2005) Educational production in Europe. *Economic Policy*, 20: 446–504.

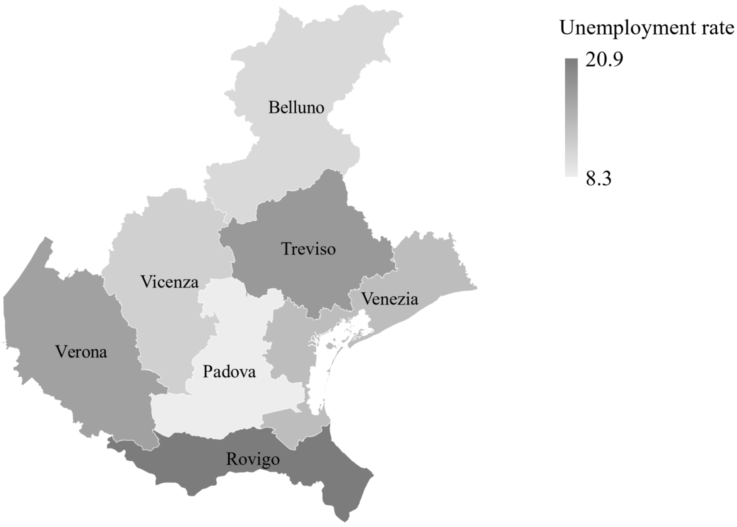
Appendix

Figure A1: Local Unemployment Rate by Veneto Province

Panel *a.* Boys

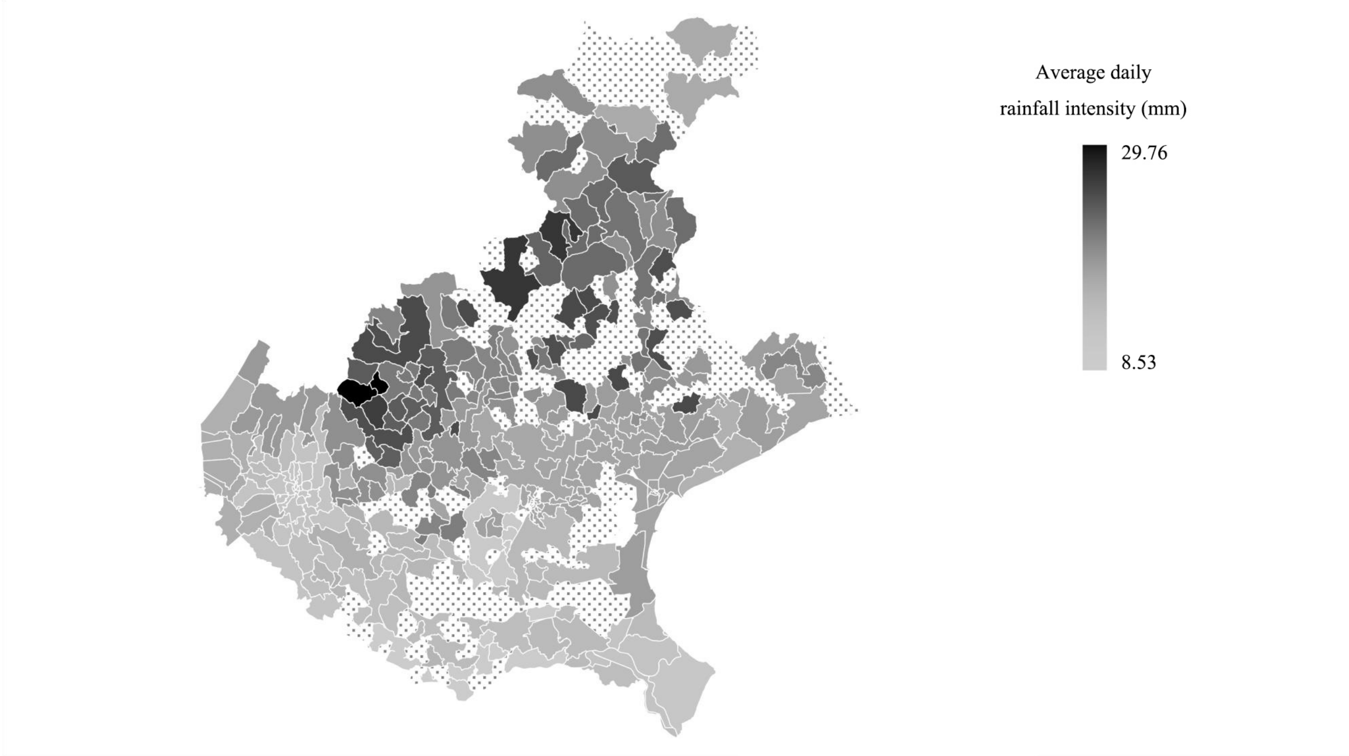


Panel *b.* Girls



Source: ISTAT (2008).

Figure A2: Daily Rainfall Intensity (millimiters) by ZIP Codes in Veneto Region



Source: Regional Agency for the Prevention and Environmental Protection of Veneto (September 2008 - June 2009).
Notes: Dotted areas refer to ZIP Codes for which we do not have students in our sample of analysis.

Table A1: Descriptive Statistics

	Boys		Girls	
	Mean	Std.Dev.	Mean	Std.Dev.
Schooling Outcome				
Math Point Average (MPA)	6.489	1.316	6.658	1.236
Effort				
Study hours	1.718	1.111	2.863	1.11
Individual and family characteristics				
Religious	0.052	0.221	0.055	0.228
Free time: sport	0.371	0.483	0.198	0.399
Free time: PS and Internet	0.122	0.327	0.040	0.195
University enrolment	0.511	0.500	0.600	0.490
Career ambition	0.448	0.498	0.454	0.498
Smoker	0.308	0.462	0.260	0.439
Patient	0.356	0.479	0.330	0.470
Risk lover	0.178	0.383	0.216	0.411
Mother: high ed.	0.165	0.371	0.132	0.339
Mother: teacher	0.053	0.225	0.046	0.209
Father: high ed.	0.161	0.368	0.151	0.358
Divorced parents	0.054	0.226	0.065	0.246
Number of siblings	1.149	0.862	1.216	0.958
Family inc.: 2nd quint.	0.186	0.389	0.211	0.408
Family inc.: 3rd quint.	0.191	0.393	0.209	0.407
Family inc.: 4th quint.	0.206	0.405	0.191	0.394
Family inc.: 5th quint.	0.242	0.429	0.171	0.376
Home owner	0.889	0.314	0.857	0.350
Urban	0.267	0.443	0.313	0.464
Parental effort	0.319	0.466	0.312	0.464
Peers' characteristics				
Share of female classmates	0.378	0.258	0.603	0.184
Classmates' average GPA	6.755	0.288	6.833	0.247
School quality				
School's FGA index	72.153	8.208	72.133	9.459
Past school performance				
Own end-of-fourth-year GPA	6.709	0.744	6.927	0.774
Observations	1184		1160	

Table A2: OLS Estimates of the Educational Production Function by Gender

	Boys		Girls	
Effort				
Study hours	-0.002	(0.033)	-0.030	(0.034)
Individual and family characteristics				
Religious	0.191	(0.146)	0.097	(0.180)
Free time: Sport	0.070	(0.072)	0.152^{*}	(0.080)
Free time: PS and Internet	0.077	(0.098)	0.155	(0.185)
University enrolment	0.180^{*}	(0.098)	-0.097	(0.083)
Career ambition	0.113^{*}	(0.058)	0.142^{*}	(0.072)
Smoker	-0.219^{***}	(0.065)	-0.087	(0.091)
Patient	0.078	(0.063)	0.043	(0.074)
Risk lover	0.129	(0.091)	-0.099	(0.079)
Mother: highly educated	-0.061	(0.134)	0.022	(0.157)
Mother: teacher	-0.085	(0.177)	-0.040	(0.189)
Father: highly educated	0.202^{*}	(0.107)	0.205^{*}	(0.115)
Divorced parents	0.030	(0.150)	-0.140	(0.141)
Number of siblings	0.092^{***}	(0.033)	0.024	(0.042)
Family income: 2nd quintile	-0.007	(0.109)	0.132	(0.109)
Family income: 3rd quintile	0.031	(0.101)	0.299^{***}	(0.086)
Family income: 4th quintile	-0.033	(0.121)	0.040	(0.098)
Family income: 5th quintile	0.079	(0.118)	0.157	(0.108)
Home owner	-0.028	(0.100)	-0.013	(0.085)
Urban	-0.078	(0.088)	-0.020	(0.060)
Parental effort	-0.086	(0.068)	-0.066	(0.079)
Peers' characteristics				
Share of female classmates	0.183	(0.250)	-0.035	(0.350)
Classmates' average GPA	-0.185	(0.207)	-0.277	(0.175)
School quality				
School's FGA index	0.001	(0.006)	-0.007	(0.005)
Past school performance				
Own end-of-fourth-year GPA	0.930^{***}	(0.045)	0.814^{***}	(0.046)
Observations	1184		1160	
Adjusted R^2	0.337		0.268	

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Robustness Check by Students' Intention to Enrol in University

	Boys			Girls		
	University enrollment		Entire	University enrollment		Entire
	Yes	No	sample	Yes	No	sample
Unemployment rate	0.036^{*}	0.029^{*}	0.031^{*}	-0.030	-0.032	-0.031
	(0.020)	(0.016)	(0.016)	(0.020)	(0.030)	(0.028)
Unemployment rate x University			0.001			-0.000
			(0.026)			(0.034)
Observations	605	579	1184	696	464	1160

Notes: Please refer to Section 4 for details about the full set of control variables. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.