Early Influences and the Gender Gap in STEM

Silvia Granato

University of Warwick

Abstract

Despite the striking reversal of the gender gap in industrialised countries in the last 40 years, women still pursue degrees in science, technology, engineering and mathematics (STEM) much less than their male peers do. I use data from a uniquely rich and largely unexplored source that combines both administrative and survey information on the population of Italian graduates to analyse the determinants of gender gaps in STEM graduation rates for Italian college leaving cohorts from 2010 to 2015, with emphasis on family, cultural and school influences, as well as geographic proximity in the supply of STEM degrees. Half of the gender gap in STEM graduation is attributed to the gender difference in maths and science content of the respective high school curricula. My results indicate that in Italy the gender gap in STEM graduation has its roots in a gendered choice originating many years before. This finding suggests that the role of the influence of environmental factors – such as the family – in the different educational choices of females and males is even greater than can be estimated through this study.

Keywords: STEM fields, choice of fields of study, gender differences

JEL Classification: I20, J16

1 Introduction

During the past 40 years there has been a striking reversal of the gender gap in education in industrialised countries. Although women are currently more likely than men to hold a college degree in the vast majority of OECD countries, their choices of college major have been and persistently continue to be different from those of men. Figure 1 illustrates the percentage of females among graduates with a bachelor degree in 7 OECD countries in 2015, for all fields of education and separately for the fields of science, engineering, education and humanities. In all countries but Germany women constitute more than half of all bachelor's degree graduates and are greatly over-represented in education and humanities, but they represent only 20 to 30% of engineering graduates.

Science, technology, engineering and mathematics degrees – indicated with the acronym STEM – have been the object of increasing attention in education, economic and policy fora. During the 2017 celebration of the International Day of Women and Girls, the UN Assistant Secretary-General Lakshmi Puri stated that *"we must ensure that women's participation in innovation is not the exception, but becomes the norm"*. Several initiatives aimed at encouraging female students to undertake STEM careers have been promoted all around the world; some examples are the initiative '*Girls in Stem*' in Turkey from the Nobel Laureate in chemistry Professor Aziz Sancar and the '*Girls in ICT*' from the International Telecommunication Union. In Italy, which is the setting for the present study, the Gender Equality Department of the government launched '*Stem Month*' in 2016, showcasing a series of initiatives targeting female pupils in primary and secondary schools, with the goal of encouraging their interest in STEM subjects.

There is a widespread consensus that STEM skills are crucial to sustaining innovation and growth (Osikominu et al., 2014). However, the share of graduates in STEM majors across OECD countries in 2015 was only 23% (and the enrolment share was approximately 27%). Thus, understanding the mechanisms underlying the educational segregation of women may shed light on issues regarding the scarcity of scientists that the European Union is concerned about.

Furthermore, several studies have provided evidence that – because STEM degrees typically lead to higher-paying jobs – gender gaps in college majors translate into gender gaps in earnings later in life (Flabbi, 2012; Anelli and Peri, 2015a; Card and Payne, 2017).

In this paper, I analyse the determinants of gender gaps in STEM graduation rates for Italian college-leaving cohorts from 2010 to 2015, with an emphasis on family, cultural and school influences, as well as geographic proximity in the supply of STEM degrees. For this purpose I use data from a uniquely rich and largely unexplored source (AlmaLaurea) that combines both administrative and survey information on the population of Italian graduates.

I am able to characterise the students' pre-college education in its most relevant aspects. One aspect is the curriculum of the high school attended, which varies widely in its maths components across a large number of available tracks. Moreover, a secondary school identifier allows me to capture the influence of unobservable school characteristics, over and above differences in their official curriculum. These administrative data are supplemented by survey-based information on students' family background and their attitudes and aspirations. By exploring the role of gender preferences in shaping college major choices I contribute to the literature on the impact of gender differences in personal traits – largely documented by the experimental literature[1]– on real-life choices.

I complement the data from AlmaLaurea with information on the general attitudes, demographic composition and political orientation of Italian municipalities. This information is then used to characterise the elements of students' background that are arguably related to gender identity norms. Finally, I use administrative data on the supply of STEM degree programmes across Italian universities in order to relate students' choices of majors to the geographic distribution of the supply of STEM degrees.

I estimate an average unadjusted gender gap in STEM graduation rates of approximately 22 percentage points for 2010-2015 cohorts. The most important determinant of this difference, driving approximately half of the observed gap, is the gender difference in the maths and science content of the respective high school curricula. This difference can be traced to educational choices made at age 14, when boys are more likely than girls to enrol into high school tracks that are more intensive in maths and science. Despite differences in high school choices, girls on average complete high school with a higher final grade than boys, regardless of track. This result implies that if girls were under-

[1] See Azmat and Petrongolo (2014) for a review of this literature.

performing relative to boys in maths- and science-intensive high school tracks, the gender gap in major choices would be even greater. Based on self-reported measures of students' personal traits, the attitudes of girls suggest lower competitiveness and higher altruism and social mindedness; however, these differences do not appear to play an important role in driving the gender difference in major choices. On the other hand, male and female students have, on average, very similar family and social environments – as measured by the parental and municipality characteristics. Therefore, the gender gap in the outcome cannot be explained by differences in these environments.

When this large set of characteristics is controlled for, half of the gap remains unexplained. The results from an Oaxaca decomposition show that approximately 50% of the part of the gap not explained by differences in characteristics is accounted for by a much lower probability of girls of choosing a STEM degree even conditional on having attended one of the maths- and science-intensive high school tracks. The results also suggest that family and social background features – over and above the influence they can already have on attitudes and previous choices – affect female and male college choices differently, each accounting for another 20% of the unexplained part of the STEM gap.

The remainder of the paper is organised as follows. Section 2 describes the conceptual framework and reviews the related literature; Section 3 describes the background of STEM college majors in the Italian education system. A description of the data and summary statistics are provided in Section 4. Section 5 presents and discusses the results based on the Gelbach and Oaxaca decompositions of the estimated gender gap in the choice of a STEM major. Section 6 concludes the paper.

2 The determinants of major choice

In this section I discuss the factors and mechanisms potentially shaping the gender gap in major choices in greater detail. I focus on three sets of explanations: (i) human capital factors, i.e., a student's preparation and achievement at pre-collegiate levels of education; (ii) personal factors, summarised by individuals' attitudes and aspirations for their future career; and (iii) parental and societal influence, which can in turn affect both high school choices and individuals' preferences for higher education.

2.1 Pre-college education

The choice of enrolling in a STEM university course is realistically influenced by the science and maths ability and knowledge that students would have acquired prior to choosing their major. This ability and knowledge are in turn largely determined by the high school track attended. In Italy, the first stage of education that offers a range of curricular choices is the start of high school, which follows the completion of middle school at age 14. Tracks available may be academic or vocational, and they vary widely in maths content. Within the academic system, high schools ("licei") specialise in one of the following: maths and science, humanities, modern languages or art. Within the vocational system, high schools ("istituti") offer a wide variety of tracks with specialisations in IT and technical applications, business and accounting, administration, tourism, etc. The distinction between the academic and vocational tracks was originally conceived to prepare students for higher education and middle-skill-level jobs, respectively. Following a law approved in 1969[2], students graduating from any high school have access to higher education. An important point to note is that in the Italian education system the choice of curriculum is made at the relatively early age of 14, when family influences may be stronger than they are later in life.

The existing literature has investigated whether boys and girls make systematically different choices prior to college entry. For the US, Xie and Schauman (2003) find that girls are less likely than boys to participate in science and engineering courses in high school. For Canada, Card and Payne (2017) find instead that the gender gap in the fraction of high school graduates who have taken STEM courses is small and is not the main explanation for the gender gap in STEM majors. My evidence for Italy demonstrates that girls are largely under-represented in maths-intensive high school tracks. In my final sample of college graduates, only 53% of girls have completed maths-intensive or technical high schools, in contrast to 83% of boys. The extent to which this gap maps to gender gaps in college majors depends on the explanatory power of the high school track in shaping major choices. Evidence for both the US and the UK indicates that taking maths-intensive courses in high school is a strong predictor of a later STEM major choice

[2] Law n.910 of the 11th of December 1969.

(Gottfried and Bozick, 2016; Philippis, 2017).

Secondary education may also impact major choices via specific (observable or unobservable) high school characteristics, over and above their general track. For example, Legewie and DiPrete (2014) find that, all else being equal, gender segregation in extracurricular activities have a discernible impact on the gender gap in the STEM choice in US. This evidence may be consistent with the self-selection of girls into high school with certain characteristics predictive of STEM choice, or with a differential gender impact of such characteristics.

Finally, conditional on high school choice, performance and final grades may play a role in STEM choice. STEM degrees are typically considered the most demanding ones; in a sample of higher education graduates from 14 OECD countries, Flabbi (2012) finds that science fields attracts the highest proportion of top-performing students in secondary school in both the male and female samples. Moreover, when looking at the perceived characteristics of the study programme, he finds that more than 20% of men and women regard study programmes in the scientific field as very demanding, while only approximately 10% of the respondents express the same judgement about humanities programmes of study. I find evidence that better high school grades are positively associated with later pursuing a STEM degree; this observation is interesting given that girls in my sample achieve, on average, better high school final grades than boys regardless of track.

2.2 Personality traits

Preferences are arguably an important factor in major choice. Wiswall and Zafar (2015) observed that the single largest factor in determining a student's college major is represented by preferences and tastes – i.e., how much the individual likes the subject and the job associated with it. This is even after randomly providing some students with additional information, such as earnings potential associated with the different majors.

Several recent studies have demonstrated that men and women are systematically different in some psychological attributes.[3] Females are found to be more risk averse and less willing to compete, and this could explain why they choose careers with less risk and

[3] See, for example, Booth and Nolen (2009), Gneezy et al. (2003), Niederle et al. (2013), Andreoni and Vesterlund (2001), Eckel and Grossman (1998).

competition. Moreover, women are found to be more socially minded and altruistic, which may translate into different occupational aspirations and career preferences. Such differences could be associated with differences in major choices, as majors in humanities and social sciences may be associated with a larger interest in society, while maths-intensive majors such as engineering may be associated with a more egoistic and competitive view of the world (Anelli and Peri, 2015b).

The evidence on the influence of these differences on real-life choices is not very rich and is mainly constrained by the lack of data adequately measuring personal traits. With respect to gender differences in college major choices, Zafar (2013) attributes the gender gap mostly to gender differences in preferences and tastes, particularly to men's stronger emphasis on pecuniary outcomes and women's stronger emphasis on enjoying their coursework and employment in potential jobs. My evidence is consistent with the following assumptions for females (compared to males): earnings are less important while culture is more important; career prospects count less, suggesting lower competitiveness; free time is valued more; and women are more involved in volunteering activities, which suggests greater social mindedness and altruism.

2.3 Family and social background

The seminal work of Akerlof and Kranton (2002) introduced the idea that individuals' social identity enters into their choices, and thus social incentives may explain why observed choices are at odds with economic incentives. Applying this idea to the gender gap in major choice implies that certain women with high ability may choose to exert lower effort and select less difficult majors with lower monetary returns when identity enters their choices, because it is expected from them under the prevailing gender identity norms and they internalise social expectations about their role. External influence can originate from a close environment, such as the family, or from broader social settings in which individuals live, such as the civic community.

A vast body of literature demonstrates positive correlations between parents and children in terms of economic, educational, social, and behavioural outcomes. Parents' educational achievement is important to the extent that it proxies for parents' abilities and skills, which are strong predictors of the abilities and skills of their children.[4] Several studies emphasise that the family environment is relevant for the transmission not only of skills but also of gender norms, and they document a positive correlation between the gender role attitudes of parents and children.[5] Cheng et al. (2017) provide interesting evidence of maternal role modelling for daughters' choices: they find that having the mother employed in a STEM occupation increases the probability of the child working in hard sciences. Thus, measuring aspects of the family arguably related to attitudes towards females, including the education or employment/social status of the mother relative to that of the father, is important in studies focusing on young students' choices.

In addition, the civic community in which individuals grow up can be important for the transmission of gender norms. Several studies indicate a direct relationship between attitudes towards women and the maths gender gap in a given society. For example, Guiso et al. (2008) compare gender differences in test performance across countries with different levels of gender equality and find that girls' under-performance in maths relative to boys' performance is eliminated in more gender-equal cultures. Moreover, González de San Román and de la Rica (2012) find that girls perform relatively better in both maths and reading in societies where gender equality is enhanced, and Nollenberger et al. (2016) demonstrate that the maths gender gap for each immigrant group living in a particular host country (and exposed to the same host country's laws and institutions) is explained by measures of gender equality in the parents' country of ancestry. The influence of the social environment can be particularly relevant in a context such as Italy, where there is a high degree of cultural diversity even across small communities such as municipalities.

3 STEM in the Italian context

The acronym STEM refers to a *"group of disciplines that teach the skills required for a high-tech economy"*.[6] What this means in practice, as well as how this definition relates to specific courses in higher education institutions, is a more complex matter; the defini-

[4]For an extensive review of the literature on the intergenerational transmission of education and earnings see Black and Devereux (2011).

[5]For example, Farré and Vella (2012) find that in a sample of US mothers and children, children's views about working women are affected by their mother's attitudes, which in turn influence female labour market decisions.

[6]Definition from the House of Lords 2nd Report 2012-2013 on Higher Education in STEM subjects.

tion varies across countries, and sometimes even among different bodies within the same country.

In Italy, a list of the university courses that are considered STEM is provided by the Ministry of Education (MIUR). These are the courses that correspond to groups 04 and 05 of the classification FOET (Fields of Education and Training) 1999: 'science, mathematics and computing' and 'engineering, manufacturing and construction'[7]. In table 1 I report the FOET 1999 classification in terms of both broad fields and a finer classification based on 'fields of education'. Within the two STEM groups, we can distinguish 7 fields: life sciences, physical sciences, maths & stats, engineering, manufacturing, architecture and building, and computing.

The STEM definition appears to include a fairly heterogeneous group of fields of study. I look at administrative data on students' enrolment in Italian universities in 2010 – made available by the MIUR – to analyse the gender gap in enrolment by field of study. The overall gender gap in enrolment in STEM fields in 2010 was 19 percentage points, with the average probability of enrolling in a STEM degree being 27%. When analysing the enrolment gender gap for each of the sub-fields (Figure 2), I find a relevant degree of heterogeneity.[8] Within STEM fields (panel (a)), the gender gap is more pronounced in some fields including computing and engineering, physics and earth science. By contrast, for other fields such as architecture, chemistry, and maths & stats the gap is smaller, or even reversed, as for manufacturing and life sciences. On the other hand, most non-STEM fields (panel (b)) are characterised by a positive gender gap; the exceptions are business and administration and most of the service fields.

To identify the characteristics that distinguish fields in which females are more likely to enrol from fields that are male-dominated, I use administrative data from the MIUR on the very detailed content of each of the approximately 2,500 unique undergraduate or single-cycle courses offered by Italian higher education institutions in 2010. I characterise the maths content of each course by building a *maths intensity* index, which is the proportion of university 'credits' that students have to obtain in maths-intensive subjects

[7]Geography is classified as physical science and is in group 04, but it is excluded from the STEM definition.

[8]I adopt here a further classification for the physical sciences group – namely, distinguishing physics, chemistry, and earth sciences – and for the architecture and building field – distinguishing architecture and town planning from building and civil engineering.

out of all the credits they need in order to graduate from a specific course. Across all courses classified as STEM, the average index is 0.64, while for non-STEM courses it is 0.13: STEM courses are clearly the maths-intensive ones. Figure 3, which plots the index separately for each STEM and non-STEM sub-field in panel (a) and panel (b), respectively, shows that maths intensity varies substantially across different fields. Within STEM fields, life science, chemistry and earth science are characterised by a relatively low maths content. Within non-STEM fields, business and administration, transport service and security service fields are characterised by a relatively high maths content.

The analysis of course content and of enrolment patterns points to a negative correlation between the maths intensity of a field and the gender gap in the probability of enrolling in majors in that field. Figure 4 plots the maths intensity and enrolment gender gap of the different fields of study on the x-axis and the y-axis, respectively. The majority of the STEM fields fall in the bottom right part of the graph; i.e., they are characterised by high maths content and a negative gender gap in enrolment. The opposite is true for most non-STEM fields. Within STEM fields, the ones characterised by a relatively lower gender gap in enrolment are also the ones with less maths content (for example chemistry, earth and life sciences), and the opposite is true within non-STEM fields (for example, business and administration and most service fields). The correlation between these two measures is -60%. Even at the level of more than 2,000 unique university courses, the correlation is almost -50%.

I will use the information obtained on course content to estimate the gender gap in the maths intensity of the specific course of study chosen and analyse its determinants.

4 Data and Variable Description

To analyse students' choices of major, I use data from the *AlmaLaurea Graduates' Profile*, a survey of the population of college graduates from most Italian universities interviewed upon graduation, which is made available by the research institution AlmaLaurea. I focus on students from undergraduate and single-cycle courses graduating from 2010 to 2015 in one of the 56 universities taking part in the survey for the whole period considered. A detailed description of the dataset and an analysis of its representativeness of the overall population of Italian college graduates are presented in the Data Appendix.

Not all students enrolled in universities will obtain a degree, and in this sense, the AlmaLaurea database represents only a selected sample of students. In particular, if the drop-out rate is differential between male and female students, this might result in an over- or underestimation of the real gender gap in the choice of studying a STEM subject. The direction of the bias is not clear *a priori*: female students might be more likely to be discouraged than male students because of their different attitudes towards competition, or women may be influenced by social pressures based on the belief that they are less suitable than men for such careers and may thus be more likely to drop out. It is also possible that only the most determined females enrol in STEM, such that STEM female students are less likely than males to drop out.

Enrolment data are available from the MIUR for the years since 2003, only aggregated at the university, field of study and province of residence level. I compare the graduation rates obtained from the AlmaLaurea data with data on enrolment rates in STEM fields by gender and year of enrolment. Figure 5 is a plot of the obtained graduation and enrolment rates and the gender gaps. The graph illustrates the lack of association between the drop-out rate in STEM fields and gender, indicating that the gender gap in graduation is a good proxy for the gender gap in the choices made by young students at time of enrolment. Given that the outcome analysed in this study is a rate resulting from the joint probability of enrolling in a STEM degree and of graduating with a STEM degree, the results of the analysis should be interpreted while noting that the impact of any factor on this outcome entails both the impact on the decision at the time of enrolment and the impact on subsequent decisions up to graduation.

For the purpose of my analysis, I exploit the richness of the *Graduates' Profile* survey to gain access to several pieces of information about each student's background. I am particularly interested in three groups of variables: (i) graduates' high school choices and performance, (ii) their attitudes and aspirations, and (iii) their family and social background.

Administrative variables provided by each university include the following: high school final grade; high school curriculum, which gives a useful measure both of students' preferences at earlier stages in life and of the type of skills they have at the moment of enrolling in the university; and the names of the specific high schools attended by each student, which allows me to control for the role of other high school characteristics over and above their general track.

The other variables are constructed from students' answers to the questionnaire. I measure students' attitudes and aspirations through answers to questions on the following: the motivation for the major choice, particularly whether professional or cultural factors had a greater influence on the decision; the relevance of several aspects related to their future career, including salary, career prospects, culture, stability and free time; the engagement in volunteering activities, which can be regarded as reflecting how altruistic and socially minded an individual is.

To characterise a student's family background, I draw on answers to questions about the level of education of both parents and their last occupation to proxy for socio-economic status. An interesting aspect of the survey is that it collects information on the field of study for parents with college degree. This information helps to distinguish and evaluate the importance of whether the students' mother and father have a STEM degree relative to other degrees.

4.1 Local variables from other data sources

An important piece of information for my analysis in the AlmaLaurea survey is the municipality of origin of each graduate. Universities provide both the municipality of birth and the municipality of residence at the time of enrolment. I draw on the latter to characterise a student's sociocultural background at the time of major choice. Secondary data sources are used to construct alternative indicators for society progressivism at different time periods and in different municipalities. The goal is to recover some indirect measures of gender equality in Italian society along two different dimensions: political empowerment and sexual emancipation.

To measure women's political empowerment, I use an indicator of whether the mayor is a female and the share of females in municipal councils, both taken from the *Census of Local and Regional Administrators* made available by the Italian Ministry of the Interior.

Following Braga and Checchi (2008), I use as proxies for women's sexual emancipation the municipality-specific fertility rate – calculated as the number of live births divided by the number of women between ages 15 and 49 times 1,000 – and the share of religious marriages over the total number of marriages, both obtained from the "*Atlante Statistico dei comuni*" of the Italian National Institute of Statistics (ISTAT). As women's control over their sexuality increases, the fertility rate should decrease. Civil marriages are characterised by lower gender segregation and a greater equality between partners.

I am able to build a consistent time series for the period between 2003 and 2011. In Figure 6, I plot the variables for 2010. Only 10% of the municipalities are governed by a female mayor, and panel (a) of the figure illustrates that these municipalities are concentrated in the northern part of the country. On average across all municipalities, the share of female councillors in local governments is only 20%, and as depicted in panel (b) the percentage is higher in northern municipalities. The average fertility rate is approximately 39 across all municipalities, and panel (d) shows that fertility is unexpectedly higher in northern regions than in southern regions, although the geographical pattern is not very clear and sharp. Finally, most marriages in Italy are celebrated with religious rituals: on average, the percentage of total marriages is 68%, and as shown in panel (d), the rate is higher in southern Italy.

4.2 Supply of STEM education

Students' decision to enrol in a STEM degree programme is potentially also a function of the availability of STEM courses. A student residing in a given municipality upon finishing high school faces a distribution of university courses offered in different locations across the country. The student's choice of major then depends not only on his/her preferences but also on the characteristics of this supply.

I use administrative data on higher education made available by the MIUR to measure the different factors characterising the higher education supply in Italy, and I summarise them in a single *supply index.* In particular, for each STEM and non-STEM course available, I extract the geographical location in which it is offered, the size of the university offering it and the availability of scholarships at the university.

An Italian student with a general high school degree can in principle choose from all of the available tertiary education programmes and institutions. For a specific group of majors – namely, most majors in the health group (medicine, dentistry) plus architecture and the recently established (2008) major educational science – access is limited and conditional on the successful performance on entry tests, which are managed nationally by the MIUR. For other majors, each offering institution can decide to set a limit on the number of students who can enrol each year. Unfortunately, information on the exact number of places made available by each university for each major characterised by nationally or locally managed limited access is not available. This makes it impossible to construct a precise measure of the availability of places supplied by each university for every field of study. By contrast, data on the number of students enrolled yearly in each major at different universities, which are easily accessible, give a measure of the equilibrium quantity resulting from the supply and demand for education. At best, this measure can be used as a proxy for the quantity of supply. In particular, I use data on enrolment to classify universities into 4 categories: very large (more than 40000 students enrolled), large (between 20000 and 40000 students enrolled), medium (between 10000 and 20000 students enrolled), and small (less than 10000 students enrolled).

The enrolment choice is also constrained by costs. Direct pecuniary costs depend on tuition fees and scholarship availability. In Italy, tuition fees are relatively low compared to international equivalents, they are similar across universities (except for a few private ones) and vary insignificantly across majors within a university. However, the availability of scholarships can vary substantially among different institutions: the level of scholarships awarded to eligible students depends on the availability of regional funds, which can vary greatly among regions. Typically, southern regions are characterised by lower availability of regional funds and consequently of scholarships relative to those available in northern regions. I draw on data on the percentage of scholarships awarded to eligible students to construct weights that confer higher relevance to universities in which the likelihood of receiving a scholarship is higher.

Another important aspect of the cost of choosing a given course of study is represented by the geographical proximity to the municipality where the course is offered. I calculate the linear distance from each Italian municipality to each municipality where a higher education course is offered. Based on the calculated distance, I construct a geographical proximity weight. This value is always 1 if the linear distance is 0 (the course is offered in the same municipality); for other municipalities, it is the inverse of the linear distance.

For each Italian municipality I construct an index by summing the number of courses – both overall and of STEM fields only – offered in all Italian municipalities, weighted by the following: the size of the university offering the course, the percentage of scholarships awarded to eligible students at each university, and the geographical proximity to the municipality where the course is offered.

Figure 7 is a plot of the resulting 2010 index for the overall supply and the STEM supply by municipality. The supply of STEM education is clearly correlated with the overall supply, but not perfectly. The figures show the dramatic difference in the supply of higher education between northern and southern Italy. Students residing in northern Italy clearly face a higher supply relative to students coming from southern regions, and this variation may account for differences in STEM graduation rates between students from different parts of the country. Assuming that male and female students are equally distributed across municipalities, these differences in the supply measure should be less relevant for the gender gap. However, if female and male students respond differently to supply, then this variable might account for part of the gender gap. For example, females might be less likely than males to leave the family and move – because of different preferences or social attitudes towards females' choices. This would imply that, given the same distance from a STEM course, females may be less likely to enrol in such a course.

4.3 Final Sample and Summary Statistics

The number of college graduates from 2010-2015 cohorts exiting from one of the 56 universities taking part in the AlmaLaurea survey for the entire period considered is approximately 1.1 million.

To analyse the choice of field of study, I focus on 3-year undergraduate or 5-year single cycle students, numbering approximately 790,000. I restrict the sample to students who were born in Italy and residing in Italy at graduation – excluding 4% of the observations – and who enrolled between the ages of 18 and 21 in the years from 2003 to 2011 – approximately 80% of the sample – which are the years for which I have data on the variables at the municipal level.

I merge these data with the data on municipality characteristics and the local supply of STEM programmes. For approximately 85% of the observations I have information on all the variables, so the final sample consists of 485,350 observations.

Table 2 lists summary statistics of the main variables presented separately for male and female students in the sample. Females constitute 62% of the sample, confirming that women are over-represented in the population of university graduates. As expected, the outcome variable documents a large gender gap in the probability of graduating in STEM fields, precisely 22 percentage points, which is 85% of the overall average probability of studying STEM. When looking at the maths intensity of the course chosen, I find a gender gap that is similar in magnitude: the percentage of maths-intensive subjects in courses chosen by females is, on average, 22 percentage points less than that for their male peers.

The distribution of the two samples across high school study paths shows that young girls are over-represented in the humanities track while boys mainly choose the scientific path.[9] The majority of men are tracked early on into classes with higher exposure to science and maths, and vice versa for girls. On the other hand, females always outperform males: they obtain a higher final high school grade on average regardless of the track chosen.

In terms of attitudes and aspirations, some interesting differences emerge: relative to men, women are less likely to declare that they have chosen their field of study for professional rather than cultural motivations, they are less likely to consider career prospects to be very important for their future job, and they seem to more strongly value aspects such as culture and stability of the job. Moreover, on average, female students carry out more volunteering activities than their male peers.

Furthermore, compared with males, females appear to have parents who are slightly less educated and have lower-level jobs.

The final group of variables included in the analysis are those measured in the municipality of residence in the year of enrolment at university, which are used to characterise the social background in which a student made the choice of major upon exiting from high school and the supply of higher education faced. Unsurprisingly, there is no difference

[9] The Scientific & Technical category is an indicator for having attended a 'scientific' high school offering students a maths- and science-intensive curriculum or a 'technical' high school offering specialisation in technological subjects such as IT, electronics or chemistry. The Humanities category is an indicator for having attended humanities-intensive high schools including 'classics', 'languages' and 'artistic' tracks.

between females and males in these variables. Thus, if any of these variable explains the gender gap in STEM graduation rates, this would not be due to differences in those environments but instead would stem from how the two sexes respond differently to similar environmental features.

5 Empirical Method and Results

I estimate a linear probability model for STEM major choice that takes into account human capital and personal factors, as well as family and societal influences. The specification estimated is given by:

$$y_{im\tau t} = \beta_1 F_i + X_i \beta_2 + Z_{m\tau}\beta_3 + \gamma m + \delta\tau + \eta t + u_{im\tau t} \quad [1]$$

where $y_{im\tau t}$ is an indicator for graduation in a STEM field for student i who resides, upon enrolment, in municipality m, enrols in year τ and graduates in year t; F_i is a female dummy; X_i is a vector of individual and family characteristics; and $Z_{m\tau}$ is a vector of variables measured at the municipal level at the time of college enrolment. I also estimate the same specification for the outcome of the maths intensity index for the college course of study chosen by each student.

The results from the full regression estimations are reported in tables B1 and B2 of the appendix for the probability of graduating from a STEM major and for the maths intensity of the specific course attended, respectively. The results are very similar for the two outcomes. From the estimations performed on the pooled sample of females and males (columns (1) of both tables) we observe that having attended a maths- and science-intensive high school and having obtained a higher high school final grade are positively associated with both outcomes. Measures of personal traits that are arguably related to a higher level of competitiveness – such as professional rather than cultural motivation for major choice and the high value attributed to career prospects and salary for one's future job – are positively associated with the outcomes. On the other hand, personal traits suggesting lower competitiveness and higher social mindedness and altruism – such as the high value attached to culture and free time in one's future job and participation in volunteering activities – are negatively related to the outcomes. A higher social status and a higher level of education of the two parents are associated both with a higher probability of graduating from a STEM major and with greater maths content of the college course. The association is stronger for parents with a STEM college degree and stronger for the father than for the mother. None of variables measured at the municipality of residence upon enrolment is significant in predicting the outcomes.

Given the estimate of the gender gap in the outcome $\hat{\beta}_1$, in order to identify and discuss the contributions of each of the five groups of variables – pre-college education, personal traits, family characteristics, social background and the supply of higher education – I adopt the conditional decomposition suggested by Gelbach (2016). Given the equation of the base model:

$$y_{im\tau t} = \tilde{\beta}_0 + \tilde{\beta}_1 F_i + \epsilon_{im\tau t} \qquad [2]$$

which gives the gender gap that we intend to decompose, Gelbach suggests a decomposition of the difference between the coefficients in the base model and the coefficient in the full model of equation [1], $(\hat{\tilde{\beta}}_1 - \hat{\beta}_1)$, given by the omitted variable bias formula: the difference is expressed as the product of the coefficient of each covariate in the full regression and the coefficient of a regression of the covariate on the female dummy. Thus, for each variable, we obtain a parameter measuring its contribution in explaining the gender gap, which is the female-male gap in the variable scaled by its STEM graduation/maths-intensity equation impact. Whether variation in a variable increases or reduces the gap depends on whether the covariate has a positive effect on the outcome and on whether the covariate has a higher mean for females or for males; thus, the Gelbach decomposition gives a very useful and intuitive way of interpreting the contribution of each covariate in explaining the gender gap.

Table 3 reports the results from this decomposition of the coefficients of both the gender gap in STEM graduation and of the maths intensity of the university course. In columns (1) and (4) – respectively for the two outcomes – I report results from the estimation of a model where the high school curriculum is included in two categories: technical or scientific versus humanities. The high school track here explains approximately 18%

of both outcomes. Among the other variables, differences in attitudes and in family characteristics each account for 2 to 4% of the gender gaps, while all the remaining variables together account for less than 1%.

In columns (2) and (5), I present results from a model in which I adopt a finer classification of the high school curriculum, which is the variable with the highest explanatory power. Within the humanities track, we can distinguish paths with a focus on classics, foreign languages, education or art; within the technical path, we can distinguish a group of tracks with a focus on business, tourism or agriculture (non-STEM) and another with a focus on industrial construction and preparation for surveyors (STEM). When the indicators for the 8 different high school tracks are included, this group of variables explains almost half (48%) of the gender gap in STEM graduation and almost 1/3 of the gap in maths intensity, while the role played by other groups of variables remains stable.

Next, I exploit the very detailed information on the secondary education institution attended by each student. I can distinguish approximately 5,500 different high schools attended by students in my sample. Some Italian high schools offer only one curriculum, while other larger ones can offer many different paths; thus, in the end, I have more than 11,000 school-track interactions. By including this information in my model, I am able to analyse the major choices conditioning not only on having chosen the same high school track but also on having attended the same secondary education institution. The results are presented in columns (3) and (6). Including the full set of school-track dummies leaves the results almost unchanged; thus, very little is due to differences in the characteristics of schools attended by females and males other than their official curriculum.

The results from the Gelbach decomposition of the estimated gap in major choices indicate overall that, among the observable measured characteristics, the most important determinant of the gap is the gender difference in the maths and science content of students' high school curriculum. At the age of 14, boys and girls are already making different educational choices, with boys more likely than girls to enrol in high school tracks that are more intensive in maths and science. Differences in self-reported measures of students' personal traits do not appear to play an important role in driving the gender difference in major choice. As expected, since male and female students come, on average, from very similar family and social environments, differences in those environments fail to explain the gender gap in outcomes. Approximately half of the gap remains unexplained by differences in observed measured characteristics.

5.1 Oaxaca Decomposition

The analysis based on the estimation of model [1] assumes that the coefficients of the covariates are the same for females and males. To account for the difference in returns to the various characteristics, I perform an Oaxaca decomposition of the regression results from the estimation of the model that includes the high school track in 8 categories. The male-female difference in the outcome is decomposed in a portion that is 'explained' by group differences in characteristics and the residual portion that cannot be accounted for by such differences in the determinants of the outcome. The decomposition method is implemented such that the difference in characteristics is weighted by coefficients for males, while the difference in coefficients is weighted by characteristics of females.

The results for both outcomes are presented in table 4; all the predictors included in the regressions are summarised in five groups, as done above. The overall gender gap is explained in approximately the same proportion by the difference in coefficients and the difference in characteristics (columns (1) and (4)).

Columns (2) and (5) report the endowment terms for each group of variables: these are equivalent to the terms of the Gelbach decomposition, with the difference being that the female-male difference in characteristics is weighted by the male coefficient instead of the coefficient from the estimation on the pooled sample. The results indicate that the group of variables that contributes the most to the portion of the gap due to differences in endowments is the high school curriculum. Females are less represented in schools with higher returns to STEM/course-maths-intensity and more represented in schools with lower returns to STEM/course-maths-intensity, and this accounts for approximately half of the overall gender difference in outcomes. The endowment term related to the high school performance is positive and relevant in magnitude, indicating that if males performed as well as females in high school, the gender gap in the outcomes would be even larger.

Columns (3) and (6) report the coefficient terms for each group of variables. Most of the overall difference in coefficients is driven by different returns from the high school track and final grade: females have lower returns to high school tracks that are positively related to the choice of a STEM degree or of courses with higher maths content, and lower returns to a higher high school final grade.

To better understand which factors within each group of variables are driving the results of the Oaxaca decomposition, I report in table 5 the detailed decomposition for each variable within the most relevant groups – namely, high school track, family and social background for the STEM graduation rate and high school track, family characteristics and attitudes for the maths intensity measure. For both outcomes, most of the difference in endowments accounted for by the high school track variables is driven by a much lower rate at which females attend a scientific and a technical STEM high school. On the other hand, the difference in returns to the high school track is driven only by a lower probability of choosing a STEM major conditional on having attended a scientific high school.

The female-male difference in returns to family characteristics (columns (2) and (4)) is mostly accounted for by the variables measuring parents' occupation: from the full regression results performed separately for the samples of females and males reported in the appendix we observe that having a parent – in particular, the father – employed in a liberal profession has a negative correlation with the probability of choosing a STEM degree only for males. This result could be due to the fact that the son, not the daughter, in those families is more likely to follow the profession of the father (or of the mother) which are typically non-STEM occupations, such as doctors or lawyers.

For the STEM graduation outcome, I look at details for the variables measuring students' social background: the most significant term is the difference in the coefficients of the variable measuring the share of religious marriages. The full regression results indicate that this variable is negatively correlated with the probability of choosing a STEM degree for females but positively correlated for males. This result suggests that in societies that are less gender equal – as measured by at least one of the variables characterising attitudes towards women in a municipality – the gender gap in the major choices is even higher.

For the maths intensity of the course, I examine details regarding the role of attitudes in explaining the gender gap: the most relevant variables are the importance of career prospects – valued less by females – and culture – valued more by females. Moreover, even assuming that females and males give the same value to career prospects, I find that females have lower probability of choosing a course with higher maths intensity.

5.2 Sub-sample Analysis

In this section, I investigate potential heterogeneity in the results across sub-samples defined according to the socio-economic status of the students' family.

The variable on socio-economic status is constructed based on the answers of students to questions regarding their parents' last occupation[10]. Through this step, three different social groups can be distinguished: low – parents in blue-collar jobs; medium – parents who are small business owners or low-level white-collar workers; and high – parents who are directors or owners of businesses with at least 15 workers or who are self-employed in liberal professions.

Tables 6 and 7 present results from the Gelbach and Oaxaca decompositions, respectively, of the gender gap in STEM graduation rates for the three sub-samples. It emerges that the lower the socio-economic status is, the higher the raw gender gap, ranging from 16 percentage points for students belonging to the highest social class to 26 percentage points for students belonging to families where parents are blue-collar workers. This result is mainly driven by the fact that females' probability of graduating from STEM programmes increases with social status while the opposite is true for males – as shown in table 7 that reports the STEM graduation rates by gender. This evidence may be consistent with the hypothesis that in families where the parents are employed in liberal professions the male sons tend to follow the profession of the parents, which are typically non-STEM professions.

While the gender gap in major choices declines with socio-economic status, the role of the different groups of variables in explaining the gap does not appear to differ significantly across the three sub-samples. Table 6 shows that the high school track explains half of the gap in each sub-sample, and except for high school performance, the other groups of variables always have negligible roles. The results from the Oaxaca decomposition,

[10] Following Schizzerotto (1994), the social class of the family refers to the highest between the two parents.

presented in table 7, are also fairly homogeneous across the different sub-samples: most of the unexplained portion of the gender gap is accounted for by lower returns of the high school track and performance for females.

The role of the high school experience as a main determinant of the different college choices of males and females is remarkably stable across social classes. This result is not completely unexpected, considering that the Italian high school system is characterised by a completely free access, such that a high level of segregation based on socio-economic status is not expected.

6 Conclusions

Despite the striking reversal of the gender gap in education in industrialised countries in the past 40 years, women pursue STEM degrees much less than their male peers do.

This paper assesses the relative importance of various explanations for the gender gap in STEM graduation rates for Italian college graduates. The major choices of students graduating from 2010 to 2015 are studied by exploiting a uniquely rich dataset obtained from the inter-university consortium AlmaLaurea. This dataset allows the measurement of students' high school experience, their attitudes and aspirations, and their family background. It is complemented with information on Italian municipalities from which I obtain measures of a student's sociocultural background characteristics, and with data on the local supply of degree programmes.

I evaluate the competing role of the different groups of variables and find that students' high school experience explains up to half of the gender gap in STEM graduation rates. Most of this is related to educational choices undertaken at an earlier stage, when young students choose between maths-intensive or humanities-oriented high school tracks. Young girls are less likely to choose tracks with a focus on maths and technical skills; this tends to refer, in particular, to the scientific 'Liceo' and the technical 'Istituto' with a focus on industrial construction and preparation for surveyors, which are the fields that ensure the highest returns to STEM enrolment in college. Even conditional on the high school track choice, a relevant role is played by the different influences of the family and social backgrounds on the decisions of females and males. Furthermore, my evidence demonstrates that females have attitudes suggesting lower competitiveness and higher altruism and social mindedness, which are negatively associated with the choice of a STEM degree, although these differences do not play a substantial role in explaining the gap in major choice.

By showing that high school track choices explain a large portion of the gender gap in STEM graduation, my results indicate that in Italy this issue has its roots in a gendered choice that has already taken place many years before. This finding suggests that the role of the influence of environmental factors – such as the family – in the different educational choices of females and males is even greater than can be estimated through this study.

These results have important policy implications. The findings indicate that effective interventions aimed at increasing girls' interests in science and technology should be implemented at an early stage, even in middle school, because the decision made by girls at 14 years of age will determine to a large extent their future education path and, consequently, their career and wage.

References

Akerlof, G. A. and Kranton, R. E. (2002). Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 40(4):1167–1201.

Andreoni, J. and Vesterlund, L. (2001). Which is the fair sex? gender differences in altruism. *The Quarterly Journal of Economics*, 116(1):293–312.

Anelli, M. and Peri, G. (2015a). Gender gap in italy: The role of college majors. In *Unexplored Dimensions of Discrimination*, pages 79–109. Oxford University Press.

Anelli, M. and Peri, G. (2015b). More unexplored dimensions of gender gap and college choice: Attitudes, choice of partner and peer/teacher effects in school. In *Unexplored Dimensions of Discrimination*, pages 110–128. Oxford University Press.

Azmat, G. and Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, 30:32–40.

Black, S. and Devereux, P. (2011). Recent developments in intergenerational mobility. volume 4B, chapter 16, pages 1487–1541. Elsevier, 1 edition.

Booth, A. and Nolen, P. (2009). Gender differences in risk behaviour: Does nurture matter? IZA Discussion Papers 4026, Institute for the Study of Labor (IZA).

Braga, M. and Checchi, D. (2008). Closing the Gender Gap? Life Competences and Social Environment. *Rivista di Politica Economica*, 98(5):155–198.

Card, D. and Payne, A. A. (2017). High school choices and the gender gap in STEM. Technical report.

Cheng, A., Kopotic, K., and Zamarro, G. (2017). Can parents' growth mindset and role modelling address stem gender gaps? Technical report, EDRE Working Paper No. 2017-07.

Eckel, C. and Grossman, P. (1998). Are women less selfish than men? evidence from dictator experiments. *Economic Journal*, 108(448):726–35.

Farré, L. and Vella, F. (2012). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica*, 80(318):219–247.

Flabbi, L. (2012). Gender Differences in Education, Career Choices and Labor Market Outcomes on a Sample of OECD Countries. *World Development Report 2012*.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal of Labor Economics*, 34(2):509–543.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

González de San Román, A. and de la Rica, S. (2012). Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes. IZA Discussion Papers 6338, Institute for the Study of Labor (IZA).

Gottfried, M. A. and Bozick, R. (2016). Supporting the STEM pipeline: Linking applied STEM course-taking in high school to declaring a STEM major in college. *Education Finance and Policy*, 11(2):177–202.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). DIVERSITY: Culture, gender, and math. *Science*, 320(5880):1164–1165.

Legewie, J. and DiPrete, T. (2014). Pathways to Science and Engineering Bachelor's Degrees for Men and Women. *Sociological Science*, 1(February):41–48.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness. *Management Science*, 59(1):1–16.

Nollenberger, N., Rodríguez-Planas, N., and Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5):257–61.

OECD (2015). Graduates by field. Online; accessed February 2018.

Osikominu, A., Grossmann, V., and Osterfeld, M. (2014). Are sociocultural factors important for studying a science university major? Annual conference 2014 (hamburg):

Evidence-based economic policy, Verein für Socialpolitik / German Economic Association.

Philippis, M. D. (2017). STEM graduates and secondary school curriculum: does early exposure to science matter? Temi di discussione (Economic working papers) 1107, Bank of Italy, Economic Research and International Relations Area.

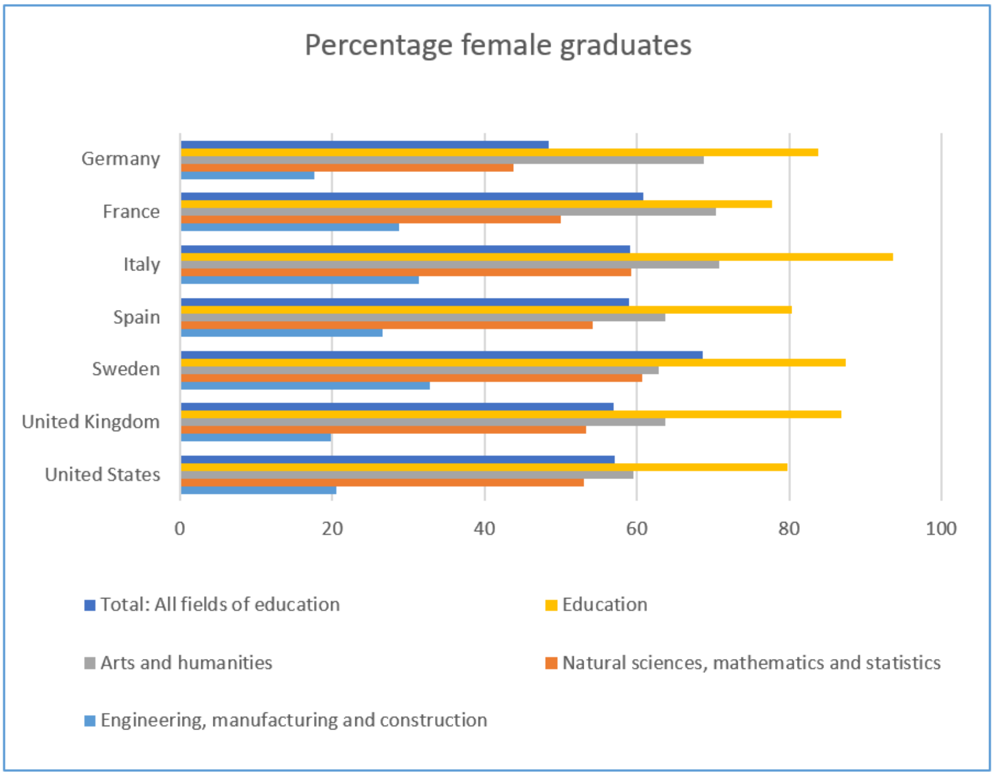
Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *Review of Economic Studies*, 82(2):791–824.

Xie, Y. and Schauman, K. A. (2003). *Women in science: Career processes and outcomes*. Cambridge: Harvard University Press.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595.

Figures and Tables

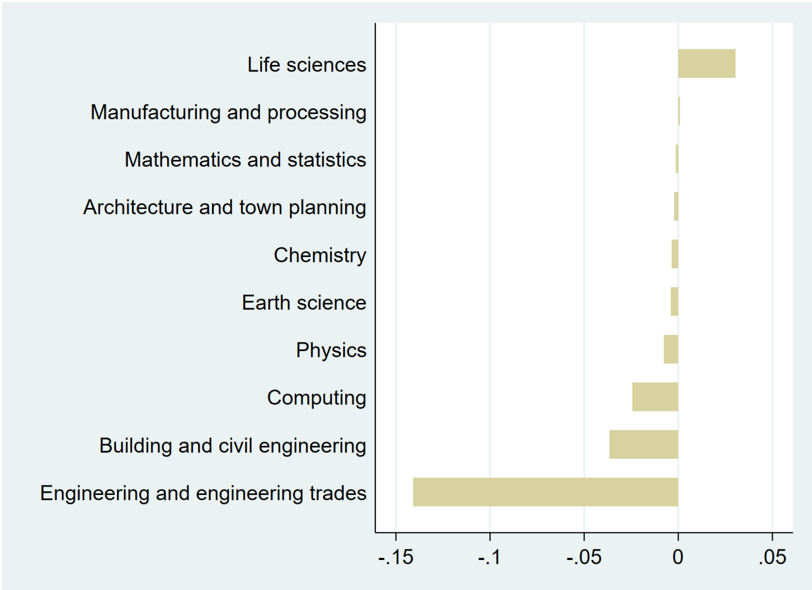
Figure 1: Gender differences in fields of study



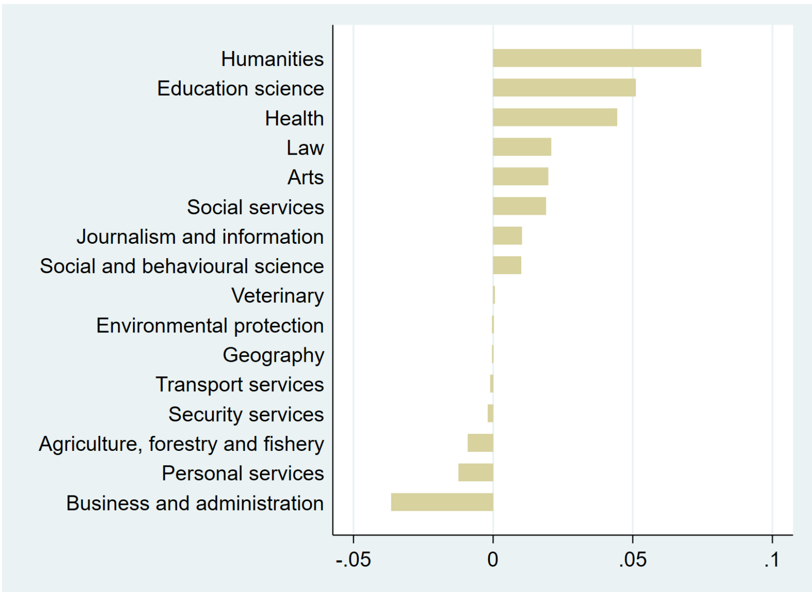
Source: OECD (2015)

Figure 2: Enrolment gender gap in fields of study

(a) STEM fields



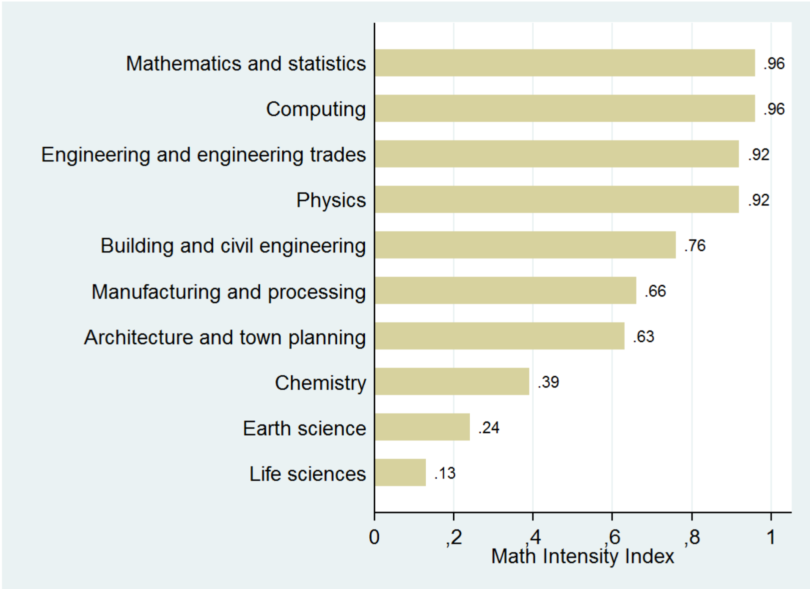
(b) non-STEM fields



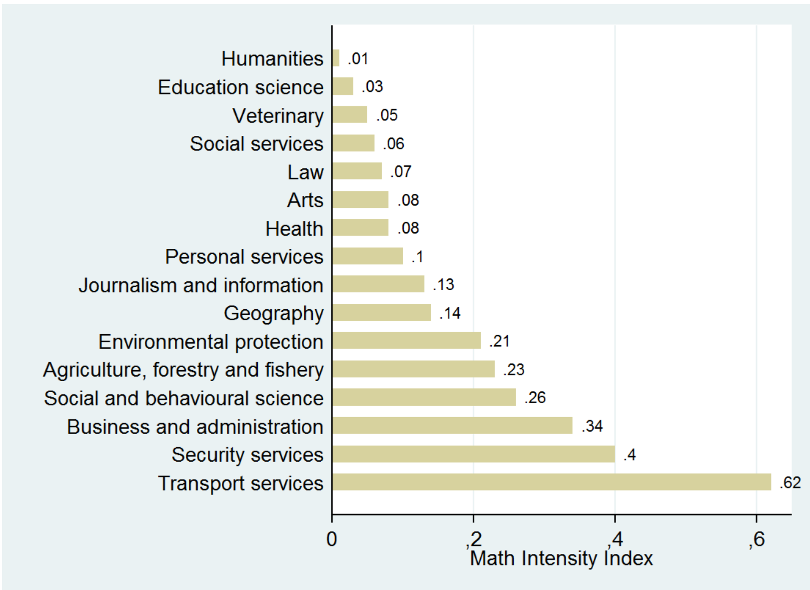
Notes: The figure plots the average female-male difference in enrolment probabilities for each group of university fields of study according to the FOET 1999 definition. Data are made available by the MIUR and are relative to the 2010/2011 academic year.

Figure 3: Maths intensity of fields of study

(a) STEM fields

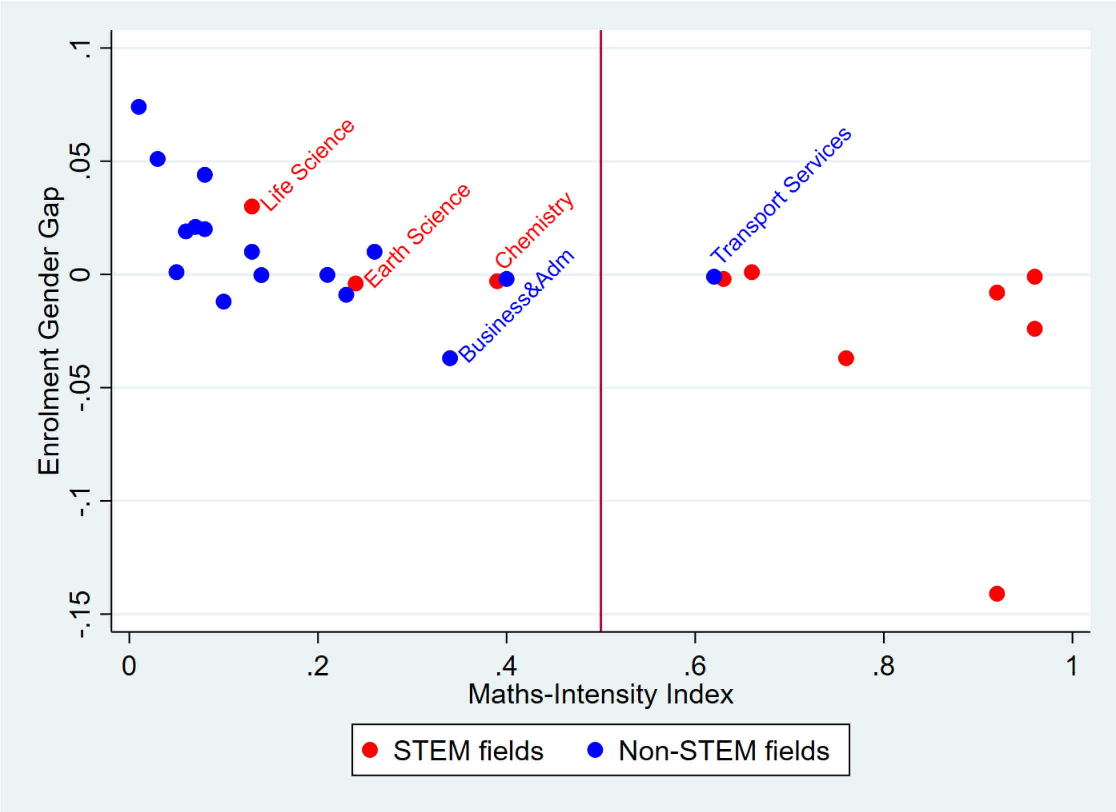


(b) non-STEM fields



Notes: The maths-intensity index is calculated as the percentage of college credits related to maths-intensive subjects out of the total credits for each field of study, averaged across all courses in a given field. Data are relative to the courses offered in the academic year 2010/2011.

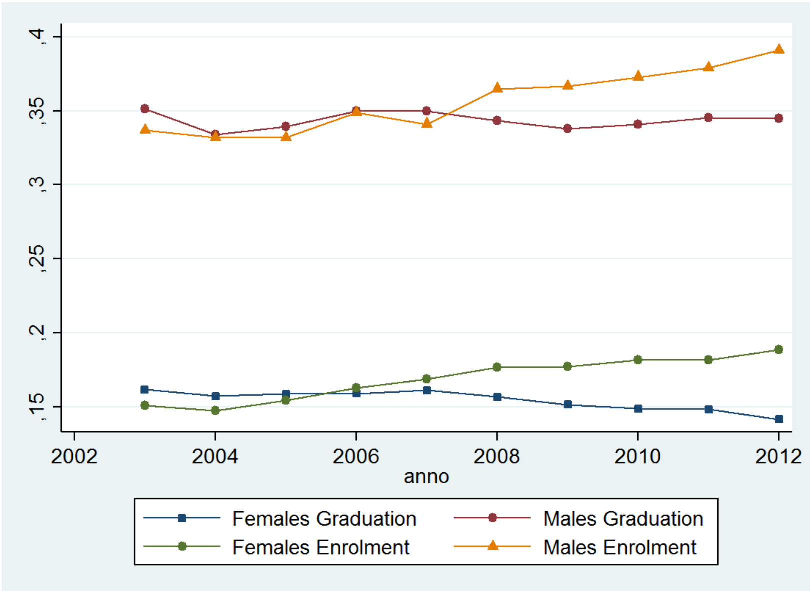
Figure 4: Enrolment gender gap and maths intensity by fields of education



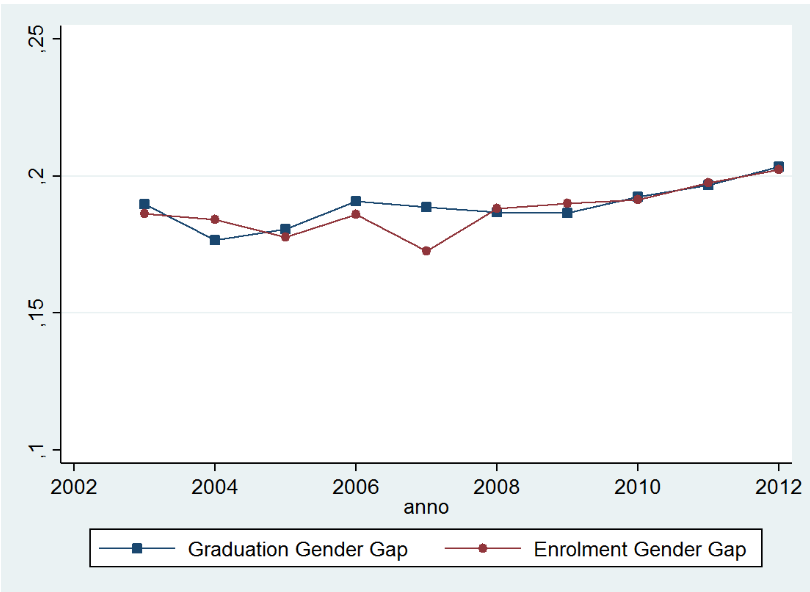
Notes: Each observation is a field of study. The average maths intensity across all courses in a given field is represented on the x-axis, while the y-axis shows the female-male difference in the probability of enrolling in each field.

Figure 5: Enrolment and graduation rates in STEM fields

(a)



(b)



Notes: Enrolment rates (number of students enrolled in STEM fields as a percentage of the total number of students enrolled) are obtained from MIUR data for students enrolled in an undergraduate or single-cycle master's degree between 2003 and 2012 in universities taking part in the AlmaLaurea survey from 2010. Graduation rates (number of students graduated from STEM fields as a percentage of the total number of graduates) are obtained from AlmaLaurea data for students who graduated from an undergraduate or single-cycle master's degree programme between 2010 and 2015 and who enrolled between 2003 and 2012, from universities taking part in the AlmaLaurea survey from 2010.

Figure 6: Municipal variables

(a) Female Mayors

(b) Share of Female Councillors

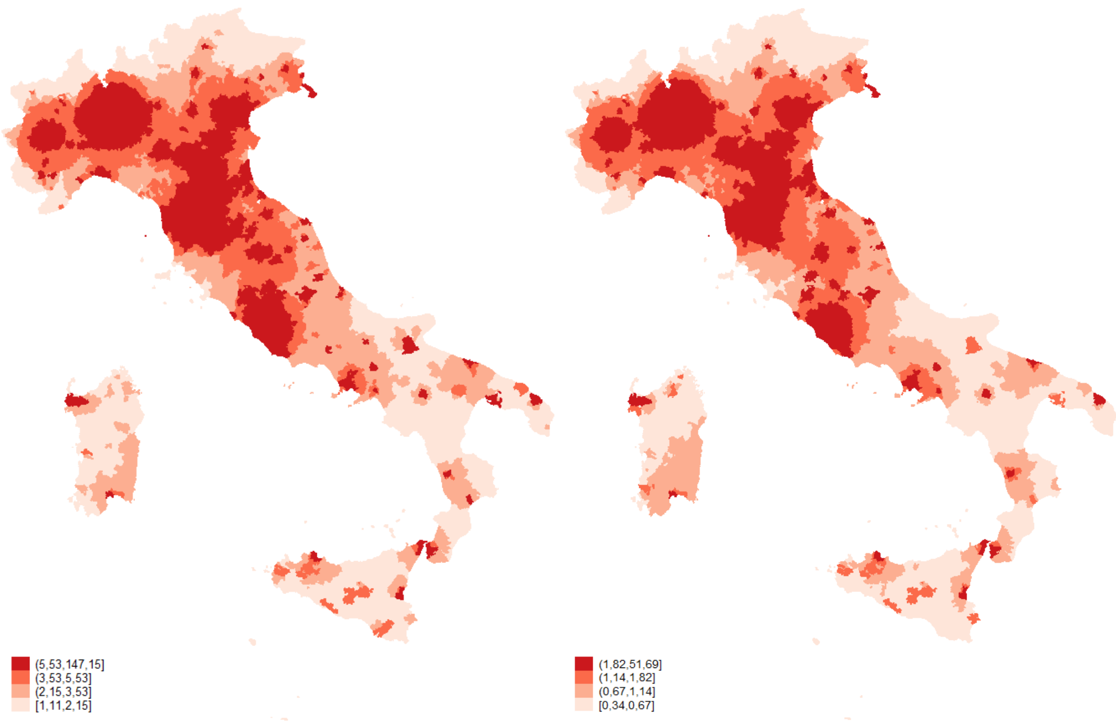
(c) Fertility Rate

(d) Share of Religious Marriages

Notes: All variables are measured in 2010. Panel (a) shows in red the municipalities governed by a female mayor, and panel (b) plots the share of female councillors in the local government at the municipal level. Both variables are obtained from data on local administrators from the Italian Ministry of the Interior. Panels (c) and (d) plot respectively the fertility rate – i.e., the ratio of the number of live births to the number of females aged 15-49 (times 1,000) – and the percentage of religious marriages, both obtained from the ISTAT *Atlante Statistico dei Comuni*.

Figure 7: Supply index

(a) All Courses (b) STEM courses



Notes: The two panels plot the index of supply in 2010, obtained for each municipality by summing the number of all/STEM-only courses offered in all other municipalities, weighted by the linear distance, the size of the university offering the course and the percentage of scholarships awarded by each university.

Table 1: FOET 1999 Classification

Broad fields	Fields of Education
1. Education	Teacher training and education science
2. Humanities and Arts	Arts
	Humanities
3. Social sciences, business and law	Social and behavioural science
	Journalism and information
	Business and administration
	Law
4. Science, Mathematics and Computing	**Life sciences**
	Physical sciences
	Mathematics and Statistics
	Computing
5. Engineering, Manufacturing and Construction	**Engineering and engineering trades**
	Manufacturing and processing
	Architecture and building
6. Agriculture	Agriculture, forestry and fishery
	Veterinary
7. Health and Welfare	Health
	Social services
8. Services	Personal services
	Transport services
	Environmental protection
	Security services

Notes: Source: Fields of Training Manual, European Centre for the Development of Vocational Training 1999

Table 2: Summary Statistics

Variables	Males		Females	
Observations	184,293		301,057	
	Mean	sd	Mean	sd
Stem	0.39	0.49	0.17	0.38
Maths intensity	0.41	0.35	0.20	0.25
High School:				
Humanities	0.18	0.38	0.47	0.50
Scientific & Technical	0.83	0.38	0.53	0.50
Final grade	80.7	12.4	83.7	12
Attitudes				
Enrolment motivation (professional)	0.12	0.33	0.08	0.28
Salary very important	0.57	0.50	0.57	0.50
Career prospects very important	0.66	0.47	0.61	0.49
Stability very important	0.65	0.48	0.75	0.43
Culture very important	0.38	0.49	0.46	0.50
Free time very important	0.26	0.44	0.26	0.44
Volunteering activities	0.21	0.41	0.25	0.43
Family Characteristics				
Father education:				
Less than HS	0.29	0.46	0.38	0.48
HS	0.46	0.50	0.44	0.50
College non STEM	0.17	0.37	0.13	0.34
College Science	0.02	0.14	0.02	0.12
College Engineering	0.06	0.23	0.04	0.20
Mother education:				
Less than HS	0.27	0.45	0.35	0.48
HS	0.51	0.50	0.48	0.50
College non STEM	0.18	0.38	0.14	0.35
College Science	0.03	0.18	0.02	0.15
College Engineering	0.01	0.09	0.01	0.08
Father last occupation:				
Blue collar (or never worked)	0.27	0.44	0.31	0.46
Self employed/small business owner	0.19	0.39	0.22	0.42
White collar	0.30	0.46	0.27	0.45
Liberal professions/entrepreneur	0.24	0.43	0.19	0.40
Mother last occupation:				
Housewife	0.23	0.42	0.26	0.44
Blue collar	0.28	0.45	0.29	0.45
Self employed/small business owner	0.10	0.30	0.11	0.31
White collar	0.32	0.47	0.29	0.45
Liberal professions/entrepreneur	0.07	0.26	0.06	0.24
Municipality Characteristics				
Fertility Rate	39.23	7.19	39.08	7.47
Religious marriages share	0.63	0.19	0.64	0.19
Female mayor	0.08	0.27	0.08	0.27
Share female councillors	0.14	0.10	0.14	0.10
Supply of STEM courses	7.8	16.0	6.9	15.0
Supply of university courses	24.5	49.7	21.7	46.6

Notes: Sample includes 3-year undergraduate or 5-year single-cycle students who enrolled between 2003 and 2011.

Table 3: Gelbach Coefficient Decomposition

Outcome:	STEM			Maths intensity		
	(1)	(2)	(3)	(4)	(5)	(6)
Estimated STEM gender gap	-0.219***			-0.218***		
	(0.00333)			(0.00228)		
HS curriculum:						
2 categories	-0.0455***			-0.0411***		
	(0.000386)			(0.000348)		
8 categories		-0.0982***			-0.0681***	
		(0.00188)			(0.00128)	
High school fixed effects			-0.103***			-0.0753***
			(0.00206)			(0.00136)
HS performance	0.0116***	0.0118***	0.0118***	0.00885***	0.00870***	0.00869***
	(0.000209)	(0.000212)	(0.000212)	(0.000160)	(0.000158)	(0.000158)
Attitudes	-0.00373***	-0.00441***	-0.00421***	-0.0114***	-0.0115***	-0.0110***
	(7.47e-05)	(8.24e-05)	(8.10e-05)	(0.000164)	(0.000165)	(0.000160)
Parents	-0.00414***	-0.00257***	-0.00231***	-0.00211***	-0.00238***	-0.00204***
	(0.000191)	(0.000175)	(0.000167)	(0.000119)	(0.000120)	(0.000113)
Municipal variables	3.30e-06	-5.87e-05***	-1.30e-05*	-6.24e-05***	-8.58e-05***	-3.95e-05***
	(9.43e-06)	(9.75e-06)	(7.47e-06)	(1.91e-05)	(2.03e-05)	(1.53e-05)
Supply	0.00131**	0.00134**	0.00152**	0.00101**	0.00101**	0.00109**
	(0.000578)	(0.000592)	(0.000674)	(0.000447)	(0.000446)	(0.000480)
Cohort fe	-8.83e-05***	0.000110***	0.000118***	0.000827***	0.00104***	0.00107***
	(1.40e-05)	(1.49e-05)	(1.47e-05)	(4.29e-05)	(5.35e-05)	(5.48e-05)
Municipality FE	-0.00281***	-0.00243***	-0.00209**	-0.00303***	-0.00281***	-0.00201***
	(0.000823)	(0.000800)	(0.000947)	(0.000623)	(0.000626)	(0.000736)
Full regression coefficient	-0.176***	-0.125***	-0.121***	-0.171***	-0.144***	-0.138***
	(0.00400)	(0.00239)	(0.00237)	(0.00253)	(0.00173)	(0.00170)
Observations	485,350	485,350	485,350	485,350	485,350	485,350
R squared	0.143	0.203	0.244	0.235	0.260	0.304

Notes: Decompositions of the gender gap in STEM graduation rate/maths intensity of university courses based on Gelbach (2016). The sample consists of college graduates who enrolled between 2003 and 2010 and graduated between 2010 and 2015. The dependent variable is a dummy equal to 1 if the individual graduated from a STEM field in columns (1)-(3) and the maths intensity of the course of study in columns (4)-(6). Each regression includes the survey year, year of graduation and municipality of residence fixed effects. The other variables are defined as follows. High school curriculum: 2 dummies for scientific/technical versus humanities in columns (1) and (4); 8 dummies for classics, education, languages, arts, technical non-STEM, technical STEM, science, and professional high school track in columns (2) and (5); more than 11,000 identifiers for secondary institution and track attended in columns (3) and (6). High school performance: 3 dummies for the intervals 60-85, 85-95, and 95-100. Attitudes: dummy=1 if the motivation to enrol in a course of study is professional versus cultural; dummies=1 if salary/career prospects/stability/culture/free time is very important versus slightly or not important in a future job; dummy=1 if engaged in volunteering activities. Parent characteristics: 5 dummies for father/mother's level of education (less than high school, high school, college non-STEM, college STEM science, and college STEM engineering); 4 dummies for father's last occupation (never worked or blue collar, small business man, white collar, liberal professions); and 5 dummies for mother's last occupation (housewife, blue collar, small business woman, white collar, liberal professions). Municipal variables: all variables measured in the municipality of residence in the year of university enrolment: dummy=1 if the mayor is female, share of female councillors, fertility rate, and share of religious marriages. Supply: indexes measuring the supply of STEM or overall university courses in the year of enrolment.

Table 4: Oaxaca Decomposition

Outcome:	STEM			Maths intensity		
	(1)	(2)	(3)	(4)	(5)	(6)
	Overall	Endowments	Coefficients	Overall	Endowments	Coefficients
Females	0.173***			0.195***		
	(0.00252)			(0.00189)		
Males	0.392***			0.413***		
	(0.00262)			(0.00210)		
Gender Gap	-0.219***			-0.218***		
	(0.00269)			(0.00188)		
Endowments	-0.0987***			-0.0853***		
	(0.00204)			(0.00146)		
Coefficients	-0.121***			-0.132***		
	(0.00234)			(0.00183)		
High School Track		-0.109***	-0.0477***		-0.0831***	-0.0405***
		(0.00208)	(0.00311)		(0.00148)	(0.00228)
High School performance		0.0177***	-0.0553***		0.0128***	-0.0389***
		(0.000442)	(0.00131)		(0.000322)	(0.000922)
Attitudes		-0.00594***	-0.000470		-0.0137***	0.00981***
		(0.000440)	(0.00257)		(0.000369)	(0.00185)
Family Characteristics		-0.00104***	0.0196***		-0.00131***	0.0107***
		(0.000359)	(0.00343)		(0.000298)	(0.00248)
Municipal Variables		4.84e-06	-0.0284*		4.87e-06	-0.00276
		(8.56e-06)	(0.0157)		(8.51e-06)	(0.0112)
Supply indexes		-3.72e-05	0.00545**		-3.80e-05	0.00509***
		(4.58e-05)	(0.00262)		(4.55e-05)	(0.00189)
Constant			-0.0137			-0.0758***
			(0.0171)			(0.0121)
Observations	485,350	485,350	485,350	485,350	485,350	485,350

Notes: Oaxaca decompositions of the gender gap in STEM graduation rate/maths intensity of university courses. The sample consists of college graduates who enrolled between 2003 and 2010 and graduated between 2010 and 2015. The dependent variable is a dummy equal to 1 if the individual graduated from a STEM field in columns (1)-(3) and the maths intensity of the course of study in columns (4)-(6). Each regression includes the survey year, year of graduation and municipality of residence fixed effects. The other variables are defined as in table 3.

Table 5: Detailed Oaxaca Decomposition

Outcome:		STEM				Maths intensity	
		(1)	(2)			(3)	(4)
VARIABLES		Endowments	Coefficients	VARIABLES		Endowments	Coefficients
High School Track	Overall	-0.109***	-0.0477***	High School Track	Overall	-0.0831***	-0.0405***
		(0.00208)	(0.00311)			(0.00148)	(0.00228)
	Education	-0.00555***	-0.00178*		Education	-0.00408***	-0.000618
		(0.000940)	(0.00102)			(0.000726)	(0.000755)
	Languages	-0.00433***	-0.00166**		Languages	-0.00303***	0.000279
		(0.000723)	(0.000844)			(0.000622)	(0.000693)
	Arts	0.00446***	-0.00311***		Arts	0.00233***	-0.00116***
		(0.000258)	(0.000330)			(0.000144)	(0.000200)
	Technical non STEM	-0.000941***	-0.00115*		Technical non STEM	0.00150***	-0.00121***
		(0.000123)	(0.000610)			(0.000167)	(0.000450)
	Technical STEM	-0.0651***	-0.00187***		Technical STEM	-0.0490***	-0.00213***
		(0.00276)	(0.000254)			(0.00207)	(0.000247)
	Science	-0.0379***	-0.0378***		Science	-0.0308***	-0.0357***
		(0.00160)	(0.00159)			(0.00128)	(0.00108)
	Professional	5.08e-06	-0.000328**		Professional	-1.11e-05	1.13e-05
		(7.61e-06)	(0.000161)			(8.14e-06)	(0.000108)
Family Characteristics	Overall	-0.00104***	0.0196***	Family Characteristics	Overall	-0.00131***	0.0107***
		(0.000359)	(0.00343)			(0.000298)	(0.00248)
	Parents education	-0.00167***	0.00613**		Parents education	-0.00153***	0.00312
		(0.000395)	(0.00279)			(0.000320)	(0.00206)
	Parents last occupation	0.000629***	0.0134***		Parents last occupation	0.000219*	0.00762***
		(0.000167)	(0.00320)			(0.000127)	(0.00223)
Municipal Variables	Overall	4.84e-06	-0.0284*	Attitudes	Overall	-0.0137***	0.00981***
		(8.56e-06)	(0.0157)			(0.000369)	(0.00185)
	Female mayor	1.08e-06	0.000112		Enrolment motivation (professional)	-0.00316***	-0.000198
		(5.23e-06)	(0.000753)			(0.000122)	(0.000245)
	Share female councillors	4.72e-07	0.00235		Salary very important	-9.87e-06	1.99e-05
		(2.36e-06)	(0.00325)			(8.62e-06)	(0.00137)
	Fertility rate	1.35e-06	-0.00842		Career prospects very important	-0.00317***	-0.0127***
		(4.17e-06)	(0.0115)			(0.000175)	(0.00134)
	Share of religious marriages	1.94e-06	-0.0225**		Stability very important	-0.000852***	0.00104
		(4.97e-06)	(0.00974)			(0.000175)	(0.00179)
					Culture very important	-0.00512***	0.0136***
						(0.000200)	(0.00109)
					Free time very important	-7.18e-05*	0.00482***
						(4.28e-05)	(0.000571)
					Volunteering activities	-0.00129***	0.00323***
						(9.05e-05)	(0.000557)

Notes: Details of the Oaxaca decomposition results presented in table 4. The table presents in columns (1) and (2) the endowment and coefficient terms of the gender gap in STEM graduation for the different variables within the groups: high school track (8 categories), family characteristics (parents' education and parents' last occupation), and municipal variables (female mayor, share of female councillors, fertility rate, share of religious marriages). In columns (3) and (4) the table presents the endowment and coefficient terms of the gender gap in maths intensity of the course of study for the different variables within the groups: high school track and family characteristics as in the other columns, and attitudes (enrolment motivation, importance of salary/career/stability/culture/free time for future jobs, involvement in volunteering activities.)

Table 6: Gelbach Decomposition by Socio-economic Status

Socio-economic status:	High			Medium			Low		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimated STEM gender gap	-0.158***			-0.229***			-0.264***		
	(0.00416)			(0.00312)			(0.00337)		
HS curriculum									
	-0.0424***			-0.0485***			-0.0469***		
	(0.000539)			(0.000451)			(0.000463)		
		-0.0655***			-0.102***			-0.122***	
		(0.00176)			(0.00156)			(0.00158)	
			-0.0661***			-0.107***			-0.130***
			(0.00215)			(0.00201)			(0.00213)
HS performance	0.0125***	0.0124***	0.0123***	0.0123***	0.0125***	0.0125***	0.00771***	0.00818***	0.00826***
	(0.000316)	(0.000312)	(0.000309)	(0.000309)	(0.000314)	(0.000315)	(0.000403)	(0.000427)	(0.000432)
Attitudes	-0.00333***	-0.00370***	-0.00365***	-0.00375***	-0.00451***	-0.00433***	-0.00457***	-0.00533***	-0.00520***
	(0.000137)	(0.000145)	(0.000148)	(9.76e-05)	(0.000108)	(0.000106)	(0.000169)	(0.000187)	(0.000185)
Parents	-0.00265***	-0.00225***	-0.00209***	-0.00362***	-0.00238***	-0.00219***	-0.00244***	-0.00154***	-0.00136***
	(0.000487)	(0.000467)	(0.000454)	(0.000172)	(0.000145)	(0.000137)	(0.000114)	(9.17e-05)	(8.49e-05)
Municipal variables	-0.000105**	-0.000105***	-6.45e-05	9.13e-05***	3.33e-05**	8.54e-05***	-5.43e-06	-5.03e-05***	-2.37e-05**
	(4.28e-05)	(3.98e-05)	(4.00e-05)	(1.71e-05)	(1.42e-05)	(2.00e-05)	(2.57e-05)	(1.71e-05)	(1.12e-05)
Supply	0.000589**	0.000579**	0.000621***	0.000753**	0.000794**	0.00102**	0.000715*	0.000822*	0.000821*
	(0.000244)	(0.000244)	(0.000236)	(0.000347)	(0.000367)	(0.000470)	(0.000378)	(0.000435)	(0.000435)
Cohort fe	0.00226***	0.00253***	0.00251***	-0.000176***	9.70e-06	2.85e-05**	-0.00188***	-0.00181***	-0.00167***
	(0.000220)	(0.000243)	(0.000241)	(2.14e-05)	(1.57e-05)	(1.37e-05)	(0.000159)	(0.000152)	(0.000143)
Municipality FE	-0.00161**	-0.00136*	-0.00166	-0.00255***	-0.00220***	-0.00145*	-0.00271***	-0.00210***	-0.000272
	(0.000751)	(0.000729)	(0.00118)	(0.000642)	(0.000609)	(0.000862)	(0.000793)	(0.000778)	(0.000983)
Full regression coefficient	-0.123***	-0.100***	-0.0993***	-0.184***	-0.132***	-0.128***	-0.214***	-0.140***	-0.134***
	(0.00515)	(0.00402)	(0.00406)	(0.00378)	(0.00256)	(0.00257)	(0.00367)	(0.00311)	(0.00332)
Observations	111,210	111,210	111,210	250,944	250,944	250,944	113,606	113,606	113,606
R squared	0.161	0.196	0.253	0.153	0.215	0.265	0.192	0.264	0.333

Notes: Decompositions of the gender gap in STEM graduation based on Gelbach (2016) for three sub-samples defined according to the socio-economic status of the students' family (high/medium/low). For each sub-sample, three models with different definitions of high school tracks are estimated: 2 dummies for scientific/technical versus humanities in columns (1),(4) and (7); 8 dummies for classics, education, languages, arts, technical non-STEM, technical STEM, science, and professional high school track in columns (2),(5) and (8); more than 11,000 identifiers for the secondary institution and track attended in columns (3),(6) and (9). The other variables are defined as in table 3.

Table 7: Sub-sample Analysis: Oaxaca Decomposition

Socio-economic status:	High			Medium			Low		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Overall	Explained	Unexplained	Overall	Explained	Unexplained	Overall	Explained	Unexplained
Females	0.200***			0.176***			0.146***		
	(0.00410)			(0.00236)			(0.00198)		
Males	0.358***			0.406***			0.410***		
	(0.00368)			(0.00252)			(0.00331)		
Gender Gap	-0.158***			-0.229***			-0.264***		
	(0.00361)			(0.00279)			(0.00338)		
Endowments	-0.0590***			-0.0997***			-0.128***		
	(0.00224)			(0.00213)			(0.00337)		
Coefficients	-0.0986***			-0.130***			-0.136***		
	(0.00398)			(0.00271)			(0.00375)		
High School Track		-0.0705***	-0.0335***		-0.112***	-0.0587***		-0.131***	-0.0900***
		(0.00223)	(0.00400)		(0.00205)	(0.00469)		(0.00323)	(0.00820)
High School performance		0.0183***	-0.0462***		0.0187***	-0.0584***		0.0126***	-0.0611***
		(0.000801)	(0.00268)		(0.000632)	(0.00185)		(0.000759)	(0.00266)
Attitudes		-0.00461***	0.00248		-0.00549***	-0.00494		-0.00780***	0.00498
		(0.000788)	(0.00486)		(0.000544)	(0.00368)		(0.000780)	(0.00562)
Family Characteristics		-0.00203***	0.00245		-0.00133***	0.0126***		-0.00139***	-0.000640
		(0.000640)	(0.0142)		(0.000329)	(0.00322)		(0.000422)	(0.00280)
Municipal Variables		-6.90e-06	-0.0542		1.27e-07	-0.0470**		-9.15e-06	-0.0267
		(2.11e-05)	(0.0404)		(2.08e-05)	(0.0224)		(4.30e-05)	(0.0338)
Supply indexes		-6.84e-05	0.0135*		-2.00e-05	0.000599		1.34e-05	0.00944***
		(5.98e-05)	(0.00695)		(4.61e-05)	(0.00415)		(4.76e-05)	(0.00307)
Constant			0.0169			0.0260			0.0280
			(0.0437)			(0.0239)			(0.0354)
Observations	111,210	111,210	111,210	250,944	250,944	250,944	113,606	113,606	113,606

Notes: Oaxaca decompositions of the gender gap in STEM graduation for three sub-samples defined according to the socio-economic status of the students' family (high/medium/low).

A Data Appendix

A.1 The AlmaLaurea Dataset

AlmaLaurea is an inter-university consortium that collects data on students who graduate from the universities that are part of the consortium. Its original institutional objectives are twofold: first, to provide member academic institutions with reliable information on their students by managing a database that collects information on graduates; second, it aims at facilitating the graduates' labour market transition by managing a service that gives firms electronic access to graduates' curriculum vitae.

Data on graduates are drawn from two different sources: first, academic institutions provide official data on students' demographic information and on their university careers. The administrative variables originated from this source are: students' date of birth, municipality of birth and of residence at time of university enrolment, high school attended and final grade, year and course of enrolment in university, university GPA, date of discussion of the dissertation and graduation grade. Second, upon graduation students complete a survey providing several pieces of information, among which: family characteristics, satisfaction from the university experience, level of other skills including language and IT skills, study experiences abroad, other training experiences, intention to continue studies, and aspirations about the future career. All these variables form the dataset referred to as *Graduates' Profile.* The historical series of this survey contains data on graduates' cohorts from 2004 to 2015. In 2010 an important variable was added to the dataset, which is the municipality where students resided upon enrolment in the course of study they graduate from.

With the goal of monitoring graduates' access to the labour market, AlmaLaurea follows graduates one, three and five years after graduation. The survey is entitled *Graduates' Employment Conditions* and provides information on: graduates' employment status, time span between graduation and first job, effectiveness of the degree for finding a job, characteristics of the current job including salary, type and location of job, and satisfaction with the job.

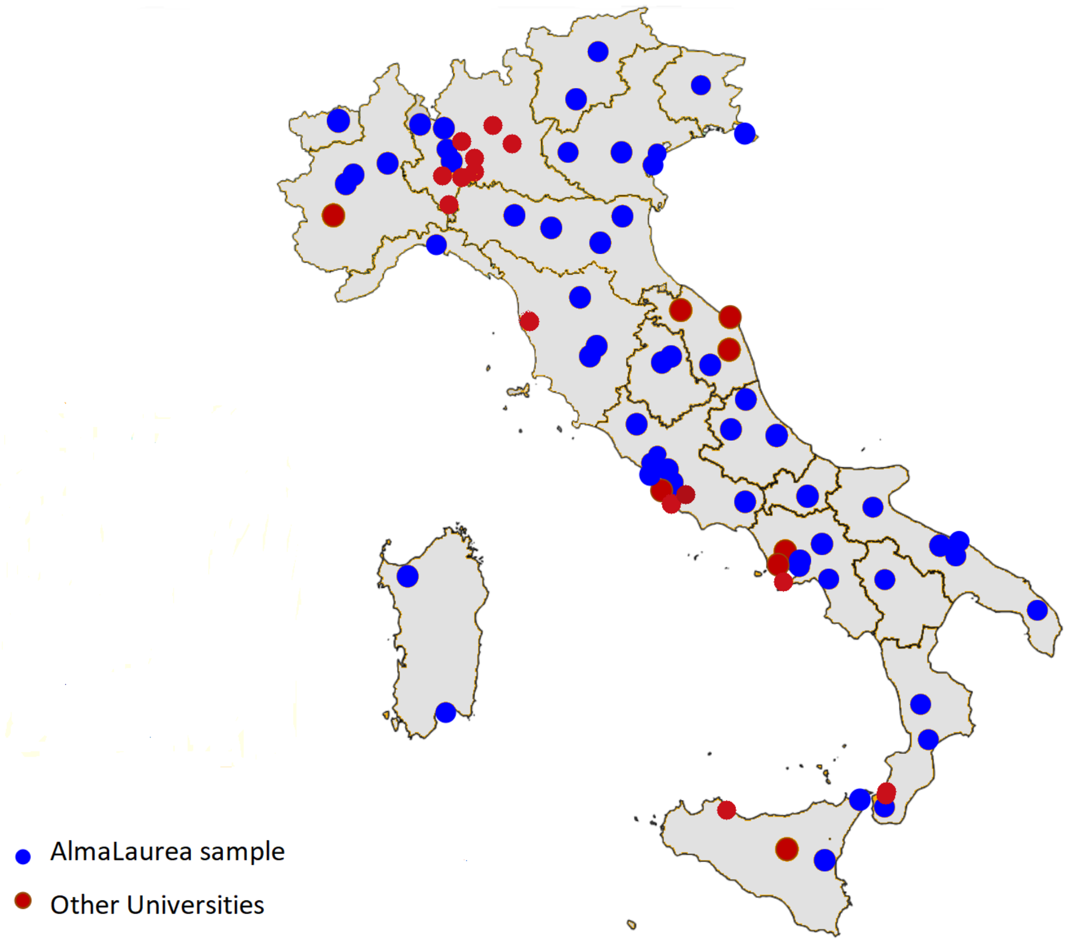
Participation in the survey from universities is voluntary: it implies the payment of a one-off membership fee and a yearly payment proportional to the total number of graduates, in exchange for the services provided by the consortium. Throughout the years more universities progressively took part in the survey. I will focus on students who graduated from 2010 to 2015 from the 56 universities surveyed every year in the period considered. The Italian higher education system in this period was composed of 89 institutions[11], including 11 long-distance-learning institutions, 3 universities for foreigners and 75 traditional universities, both public and private. Figure A1 illustrates the geographical distribution of the Italian universities (excluding the long-distance-learning institutions) highlighting those that are in the AlmaLaurea sample. There are important institutions that are not part of the sample in the period considered: namely, the two most important state universities, the technical university and the two major private universities in a major city in the north-east of the country (Milan); the biggest university in a major city in southern Italy (Naples); and a very important university in Sicily. In table A1 I report the distribution of the universities in the population and in the AlmaLaurea sample across various dimensions. It can be noticed that there are no significant differences in terms of size or field of study of the courses offered by the institutions. The AlmaLaurea sample contains no long-distance-learning institutions, while public universities are more represented.

Overall across all cohorts the AlmaLaurea sample covers approximately 65% of the population of the Italian college graduates; panel A of table A2 lists the details of the coverage by type of degree distinguishing undergraduate, single-cycle and master's degrees. Panel B reports the distribution of students across fields of study by gender in the population and in the sample, and demonstrates that the two distributions are very close.

Once a university takes part in the consortium, it provides administrative information on the universe of its graduates. Response rate to the questionnaire at graduation is very high: between 91 and 93% of students complete the survey each year. Three years after graduation the response rate is still remarkably high, ranging between 74 and 80%. In table A3 I report the response rate at graduation and three years after, by graduation cohort and type of degree.

[11]Excluding two institutions accredited respectively in 2011 and 2014.

Figure A1: Map of the Italian higher education system



Notes: The figure plots the 78 (non long-distance-learning) Italian higher education institutions existing in 2015, by geographical location and distinguishing those not surveyed by AlmaLaurea.

Table A1: AlmaLaurea Sample: Universities

Distribution of Universities (% over total)

	AlmaLaurea sample	All Universities
Size (n. students)		
<10000	41.07	43.82
10000-20000	21.43	22.47
20000-40000	25	21.35
>40000	12.5	12.36
Type		
Long-Distance-Learning	0	12.4
Private	8.93	11.2
Public	91.07	76.41
Courses offered by field		
Education	9.6	9.86
Humanities and Arts	14.9	14.08
Social sciences, business and law	16.56	18.08
Science, Maths and Computing	13.58	12.44
Engineering, Manufacturing	12.91	12.68
Agriculture	7.28	7.28
Health and Welfare	13.25	13.15
Services	11.92	12.44

Notes: Data on the population of Italian universities and number of graduates are taken from the Office of Statistics of the Italian Ministry of Education.

Table A2: AlmaLaurea Sample: Students

Panel A: Number of students by type of degree

	Population	AL sample	% of population in AL sample
Undergraduate	1,029,077	672,068	0.65
Single cycle	175,342	115,890	0.66
Master's degree	518,647	337,066	0.65
Total	1,723,066	1,125,024	0.65

Panel B: Distribution of students by gender and field of study (% over total)

Field of study	AlmaLaurea sample			All Universities		
	Males	Females	All	Males	Females	All
Education	0.8	6.0	3.9	0.8	5.8	3.7
Humanities and Arts	9.5	20.3	16.0	8.3	19.2	14.7
Social sciences, business and law	32.2	34.7	33.7	34.9	35.8	35.4
Science, Maths and Computing	10.2	7.7	8.7	9.4	7.9	8.5
Engineering and Manufacturing	29.2	9.9	17.6	29.3	10.4	18.3
Agriculture	2.7	1.6	2.0	2.6	1.6	2.0
Health and welfare	11.8	17.4	15.1	10.6	16.7	14.1
Services	3.7	2.5	3.0	4.0	2.6	3.2

Notes: Data on the population of Italian graduates are taken from the Office of Statistics of the Italian Ministry of Education.

Table A3: Response Rate by Graduation Cohort and Type of Degree

Panel A: Response rate at graduation (%)

Year of graduation	Type of degree			
	Undergraduate	Single cycle	Master	Total
2010	92	90	91	0.91
2011	94	91	92	0.93
2012	93	91	90	0.92
2013	93	92	91	0.92
2014	93	93	90	0.92
2015	93	92	91	0.92
Total	93	91	91	92

Panel B: Response rate three years after graduation (%)

Year of graduation	Type of degree			
	Undergraduate	Single cycle	Master	Total
2010	-	78	80	80
2011	-	76	76	76
2012	-	74	75	74
Total	-	76	77	77

Notes: The sample consists of all college graduates from the 56 universities surveyed by AlmaLaurea every year from 2010.

B Appendix

Table B1: Full Regressions: STEM Graduation Rate

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	-0.176***			-0.125***			-0.121***		
	(0.00400)			(0.00239)			(0.00237)		
High School track (humanities excluded):									
Scientific/Technical	0.156***	0.236***	0.127***						
	(0.00236)	(0.00329)	(0.00229)						
Education				-0.0605***	-0.0965***	-0.0682***			
				(0.00324)	(0.00789)	(0.00311)			
Languages				-0.0557***	-0.0660***	-0.0614***			
				(0.00308)	(0.00680)	(0.00298)			
Arts				0.249***	0.364***	0.218***			
				(0.00802)	(0.0141)	(0.00838)			
Technical non STEM				-0.0705***	-0.0526***	-0.0650***			
				(0.00264)	(0.00476)	(0.00297)			
Technical STEM				0.396***	0.434***	0.371***			
				(0.00424)	(0.00574)	(0.00878)			
Science				0.179***	0.246***	0.148***			
				(0.00226)	(0.00431)	(0.00247)			
Professional				-0.0274***	-0.00449	-0.0273***			
				(0.00479)	(0.00918)	(0.00519)			
School dummies							YES	YES	YES
High school final grade:									
85-95	0.0916***	0.156***	0.0506***	0.0930***	0.152***	0.0553***	0.0921***	0.152***	0.0557***
	(0.00188)	(0.00321)	(0.00197)	(0.00190)	(0.00318)	(0.00195)	(0.00196)	(0.00329)	(0.00201)
95-100	0.140***	0.231***	0.0863***	0.143***	0.228***	0.0923***	0.143***	0.231***	0.0940***
	(0.00247)	(0.00391)	(0.00237)	(0.00244)	(0.00377)	(0.00234)	(0.00265)	(0.00407)	(0.00250)
Attitudes									
Enrolment motivation (professional)	0.0190***	0.0388***	-0.00121	0.0204***	0.0373***	0.00293	0.0196***	0.0355***	0.00262
	(0.00234)	(0.00347)	(0.00264)	(0.00222)	(0.00336)	(0.00255)	(0.00224)	(0.00351)	(0.00254)
Salary very important	0.00394***	-0.00133	0.00734***	0.00507***	0.000155	0.00740***	0.00445***	-0.00118	0.00722***
	(0.00143)	(0.00275)	(0.00200)	(0.00137)	(0.00265)	(0.00192)	(0.00144)	(0.00267)	(0.00205)
Career prospects very important	0.0127***	0.0194***	0.00454**	0.0147***	0.0212***	0.00830***	0.0148***	0.0222***	0.00813***
	(0.00160)	(0.00302)	(0.00177)	(0.00155)	(0.00285)	(0.00174)	(0.00146)	(0.00282)	(0.00173)
Stability very important	0.00554***	0.0161***	-0.00308	0.00623***	0.0141***	0.000254	0.00823***	0.0161***	0.00221
	(0.00156)	(0.00250)	(0.00213)	(0.00149)	(0.00236)	(0.00209)	(0.00141)	(0.00240)	(0.00194)
Culture very important	-0.0205***	-0.0353***	-0.0107***	-0.0277***	-0.0384***	-0.0203***	-0.0282***	-0.0403***	-0.0205***
	(0.00140)	(0.00289)	(0.00141)	(0.00138)	(0.00281)	(0.00138)	(0.00135)	(0.00286)	(0.00139)
Free time very important	-0.0237***	-0.0358***	-0.0159***	-0.0214***	-0.0345***	-0.0130***	-0.0195***	-0.0309***	-0.0120***
	(0.00153)	(0.00268)	(0.00167)	(0.00151)	(0.00256)	(0.00167)	(0.00156)	(0.00274)	(0.00175)
Volunteering activities	-0.0300***	-0.0400***	-0.0255***	-0.0304***	-0.0370***	-0.0266***	-0.0301***	-0.0369***	-0.0263***
	(0.00141)	(0.00277)	(0.00168)	(0.00136)	(0.00264)	(0.00162)	(0.00133)	(0.00263)	(0.00162)

cont'd

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Father education (less than HS excluded):									
High school	0.0154***	0.00794**	0.0192***	0.00885***	0.00897***	0.00929***	0.00779***	0.00805**	0.00767***
	(0.00158)	(0.00314)	(0.00165)	(0.00147)	(0.00296)	(0.00156)	(0.00152)	(0.00314)	(0.00161)
College non STEM	-0.0264***	-0.0571***	0.00233	-0.0319***	-0.0509***	-0.0138***	-0.0311***	-0.0499***	-0.0139***
	(0.00343)	(0.00524)	(0.00332)	(0.00321)	(0.00502)	(0.00321)	(0.00292)	(0.00491)	(0.00305)
College STEM Science	0.0903***	0.0752***	0.102***	0.0779***	0.0740***	0.0806***	0.0740***	0.0681***	0.0781***
	(0.00537)	(0.00819)	(0.00651)	(0.00531)	(0.00815)	(0.00646)	(0.00508)	(0.00780)	(0.00677)
College STEM Engineering	0.152***	0.167***	0.139***	0.135***	0.156***	0.115***	0.129***	0.150***	0.110***
	(0.00442)	(0.00627)	(0.00545)	(0.00471)	(0.00671)	(0.00547)	(0.00455)	(0.00726)	(0.00525)
Mother education (less than HS excluded):									
High school	0.00876***	-0.00488	0.0151***	0.00362**	0.000380	0.00553***	0.00332**	0.00129	0.00439**
	(0.00151)	(0.00310)	(0.00168)	(0.00146)	(0.00287)	(0.00177)	(0.00148)	(0.00299)	(0.00180)
College non STEM	0.0128***	0.000646	0.0237***	0.00571**	0.00587	0.00769***	0.00465*	0.00482	0.00577**
	(0.00282)	(0.00535)	(0.00267)	(0.00262)	(0.00500)	(0.00262)	(0.00245)	(0.00487)	(0.00267)
College STEM Science	0.0906***	0.0679***	0.107***	0.0788***	0.0709***	0.0844***	0.0739***	0.0671***	0.0779***
	(0.00545)	(0.00755)	(0.00681)	(0.00534)	(0.00731)	(0.00679)	(0.00475)	(0.00693)	(0.00637)
College STEM Engineering	0.124***	0.116***	0.133***	0.112***	0.115***	0.112***	0.105***	0.107***	0.103***
	(0.0104)	(0.0146)	(0.0117)	(0.0104)	(0.0150)	(0.0116)	(0.00901)	(0.0140)	(0.0110)
Father last occupation (blue collar or never worked excluded):									
Self-employed/small businessman	0.00398**	0.000862	0.00653***	0.00424**	0.00466	0.00448**	0.00340**	0.00333	0.00403**
	(0.00179)	(0.00348)	(0.00188)	(0.00171)	(0.00330)	(0.00183)	(0.00167)	(0.00325)	(0.00188)
White collar	0.0118***	0.00863***	0.0128***	0.00890***	0.00953***	0.00797***	0.00792***	0.00869***	0.00757***
	(0.00176)	(0.00300)	(0.00217)	(0.00174)	(0.00287)	(0.00214)	(0.00175)	(0.00303)	(0.00213)
Liberal professions/white collar director/entrepreneur	0.00168	-0.0158***	0.0153***	-0.00133	-0.0131***	0.00816***	-0.00166	-0.0141***	0.00764***
	(0.00214)	(0.00375)	(0.00278)	(0.00212)	(0.00357)	(0.00274)	(0.00216)	(0.00360)	(0.00290)
Mother last occupation (housewife excluded):									
Blue collar	-0.00285	-0.00662*	-0.000423	-0.00507***	-0.00920***	-0.00292	-0.00659***	-0.0118***	-0.00420**
	(0.00179)	(0.00358)	(0.00201)	(0.00173)	(0.00343)	(0.00196)	(0.00168)	(0.00348)	(0.00197)
Self-employed/small businessman	0.000380	-0.0160***	0.0108***	-0.000774	-0.0128***	0.00656**	-0.000416	-0.0105**	0.00556**
	(0.00247)	(0.00449)	(0.00285)	(0.00234)	(0.00420)	(0.00272)	(0.00235)	(0.00428)	(0.00275)
White collar	0.00573***	-0.00811**	0.0159***	0.00198	-0.00949***	0.0100***	0.000626	-0.0106***	0.00866***
	(0.00192)	(0.00384)	(0.00242)	(0.00185)	(0.00364)	(0.00238)	(0.00183)	(0.00362)	(0.00240)
Liberal professions/white collar director/entrepreneur	-0.0109***	-0.0222***	-0.000139	-0.0134***	-0.0221***	-0.00605*	-0.0133***	-0.0201***	-0.00748**
	(0.00286)	(0.00511)	(0.00351)	(0.00281)	(0.00487)	(0.00355)	(0.00268)	(0.00520)	(0.00337)
Municipal Variables									
Female mayor	0.00519	0.00516	0.00456	0.00317	0.00361	0.00284	0.00367	0.00218	0.00496
	(0.00546)	(0.0105)	(0.00439)	(0.00527)	(0.00949)	(0.00434)	(0.00515)	(0.00887)	(0.00455)
Share female councillors	0.00325	-0.00894	0.0140	0.00514	0.00182	0.0121	0.00531	0.00540	0.00940
	(0.0170)	(0.0250)	(0.0179)	(0.0165)	(0.0235)	(0.0175)	(0.0169)	(0.0246)	(0.0175)
Fertility rate	0.000117	0.000371	-5.75e-06	0.000170	0.000421	8.24e-05	0.000128	0.000489*	3.49e-05
	(0.000150)	(0.000288)	(0.000166)	(0.000144)	(0.000273)	(0.000160)	(0.000144)	(0.000284)	(0.000163)
Share of religious marriages	0.000826	0.0258*	-0.0108	-0.00362	0.0189	-0.0142*	-5.30e-05	0.0264*	-0.0132
	(0.00766)	(0.0147)	(0.00851)	(0.00742)	(0.0141)	(0.00825)	(0.00738)	(0.0144)	(0.00832)
Supply of STEM courses	-0.00379	-0.00589***	-0.00242	-0.00410	-0.00566***	-0.00290	-0.00373	-0.00601**	-0.00279
	(0.00252)	(0.00225)	(0.00348)	(0.00253)	(0.00212)	(0.00337)	(0.00286)	(0.00240)	(0.00338)
Supply of university courses	0.000776	0.00127*	0.000503	0.000865	0.00123*	0.000608	0.000677	0.00132*	0.000467
	(0.000744)	(0.000726)	(0.00109)	(0.000748)	(0.000668)	(0.00106)	(0.000820)	(0.000692)	(0.00106)
Observations	485,350	183,588	300,787	485,350	183,588	300,787	485,350	181,294	299,321
R-squared	0.143	0.142	0.086	0.203	0.211	0.135	0.244	0.270	0.181

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: The sample consists of college graduates who were 18 between 2003 and 2010 and who graduated between 2010 and 2015. The dependent variable is a binary variable equal 1 if the individual graduated from a STEM field. Each regression includes the survey year, year of graduation and municipality of residence fixed effects.

Table B2: Full Regressions: Maths Intensity of University Courses

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	-0.171***			-0.144***			-0.138***		
	(0.00253)			(0.00173)			(0.00170)		
High School track (humanities excluded):									
Scientific/Technical	0.141***	0.207***	0.117***						
	(0.00150)	(0.00273)	(0.00141)						
Education				-0.0363***	-0.0710***	-0.0449***			
				(0.00239)	(0.00593)	(0.00226)			
Languages				-0.0257***	-0.0457***	-0.0308***			
				(0.00264)	(0.00599)	(0.00250)			
Arts				0.146***	0.191***	0.133***			
				(0.00475)	(0.00855)	(0.00509)			
Technical non STEM				0.0776***	0.0945***	0.0802***			
				(0.00262)	(0.00427)	(0.00286)			
Technical STEM				0.286***	0.329***	0.242***			
				(0.00308)	(0.00422)	(0.00606)			
Science				0.137***	0.200***	0.107***			
				(0.00164)	(0.00323)	(0.00160)			
Professional				0.00632*	0.0168**	0.0130***			
				(0.00358)	(0.00692)	(0.00367)			
School dummies									
High school final grade:									
85-95	0.0676***	0.111***	0.0402***	0.0662***	0.107***	0.0400***	0.0656***	0.107***	0.0403***
	(0.00127)	(0.00231)	(0.00128)	(0.00131)	(0.00229)	(0.00132)	(0.00135)	(0.00245)	(0.00133)
95-100	0.109***	0.168***	0.0736***	0.107***	0.164***	0.0730***	0.107***	0.166***	0.0738***
	(0.00169)	(0.00291)	(0.00160)	(0.00172)	(0.00287)	(0.00164)	(0.00190)	(0.00325)	(0.00171)
Attitudes									
Enrolment motivation (professional)	0.0833***	0.0834***	0.0808***	0.0841***	0.0833***	0.0818***	0.0822***	0.0804***	0.0815***
	(0.00190)	(0.00242)	(0.00237)	(0.00187)	(0.00237)	(0.00234)	(0.00184)	(0.00235)	(0.00234)
Salary very important	0.00396***	0.00203	0.00473***	0.00465***	0.00316*	0.00459***	0.00462***	0.00209	0.00509***
	(0.00114)	(0.00183)	(0.00153)	(0.00112)	(0.00177)	(0.00152)	(0.00116)	(0.00179)	(0.00160)
Career prospects very important	0.0452***	0.0546***	0.0362***	0.0444***	0.0538***	0.0362***	0.0428***	0.0529***	0.0343***
	(0.00118)	(0.00208)	(0.00131)	(0.00114)	(0.00199)	(0.00128)	(0.00104)	(0.00198)	(0.00125)
Stability very important	-0.00806***	-0.00630***	-0.00978***	-0.00840***	-0.00787***	-0.00861***	-0.00644***	-0.00548***	-0.00720***
	(0.00102)	(0.00177)	(0.00138)	(0.000976)	(0.00173)	(0.00133)	(0.000928)	(0.00168)	(0.00126)
Culture very important	-0.0456***	-0.0638***	-0.0336***	-0.0474***	-0.0640***	-0.0362***	-0.0462***	-0.0630***	-0.0348***
	(0.000988)	(0.00220)	(0.000922)	(0.000948)	(0.00213)	(0.000927)	(0.000932)	(0.00213)	(0.000963)
Free time very important	-0.0214***	-0.0311***	-0.0145***	-0.0212***	-0.0313***	-0.0139***	-0.0187***	-0.0276***	-0.0122***
	(0.00103)	(0.00189)	(0.00116)	(0.00102)	(0.00186)	(0.00118)	(0.00104)	(0.00194)	(0.00121)
Volunteering activities	-0.0258***	-0.0343***	-0.0216***	-0.0253***	-0.0326***	-0.0210***	-0.0247***	-0.0318***	-0.0205***
	(0.00104)	(0.00212)	(0.00103)	(0.00102)	(0.00207)	(0.00101)	(0.00103)	(0.00209)	(0.00104)

cont'd

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Father education (less than HS excluded):									
High school	0.0111***	0.00884***	0.0118***	0.0108***	0.0120***	0.0100***	0.00924***	0.0106***	0.00800***
	(0.00113)	(0.00224)	(0.00115)	(0.00108)	(0.00217)	(0.00112)	(0.00111)	(0.00231)	(0.00115)
College non STEM	-0.0206***	-0.0399***	-0.00133	-0.0174***	-0.0310***	-0.00412*	-0.0184***	-0.0329***	-0.00553***
	(0.00296)	(0.00463)	(0.00232)	(0.00259)	(0.00434)	(0.00211)	(0.00231)	(0.00426)	(0.00200)
College STEM Science	0.0413***	0.0389***	0.0440***	0.0424***	0.0445***	0.0403***	0.0387***	0.0393***	0.0378***
	(0.00450)	(0.00695)	(0.00469)	(0.00432)	(0.00681)	(0.00456)	(0.00379)	(0.00644)	(0.00438)
College STEM Engineering	0.0985***	0.108***	0.0899***	0.0972***	0.109***	0.0857***	0.0916***	0.102***	0.0808***
	(0.00331)	(0.00482)	(0.00368)	(0.00332)	(0.00506)	(0.00354)	(0.00307)	(0.00514)	(0.00343)
Mother education (less than HS excluded):									
High school	0.00228**	-0.00438**	0.00529***	0.00340***	0.00168	0.00447***	0.00312***	0.00283	0.00375***
	(0.000989)	(0.00207)	(0.00105)	(0.000963)	(0.00199)	(0.00107)	(0.000974)	(0.00206)	(0.00108)
College non STEM	0.00425**	-0.000509	0.00962***	0.00702***	0.00802**	0.00780***	0.00578***	0.00647*	0.00655***
	(0.00209)	(0.00377)	(0.00200)	(0.00182)	(0.00348)	(0.00189)	(0.00165)	(0.00336)	(0.00185)
College STEM Science	0.0524***	0.0432***	0.0589***	0.0552***	0.0520***	0.0568***	0.0505***	0.0471***	0.0529***
	(0.00394)	(0.00523)	(0.00469)	(0.00370)	(0.00508)	(0.00446)	(0.00319)	(0.00470)	(0.00416)
College STEM Engineering	0.0777***	0.0792***	0.0788***	0.0778***	0.0848***	0.0744***	0.0723***	0.0775***	0.0684***
	(0.00637)	(0.00869)	(0.00749)	(0.00604)	(0.00863)	(0.00717)	(0.00514)	(0.00833)	(0.00655)
Father last occupation (blue collar or never worked excluded):									
Self-employed/small businessman	0.0127***	0.0111***	0.0143***	0.0136***	0.0139***	0.0138***	0.0123***	0.0140***	0.0120***
	(0.00127)	(0.00253)	(0.00129)	(0.00126)	(0.00245)	(0.00131)	(0.00126)	(0.00249)	(0.00133)
White collar	0.0102***	0.0100***	0.00939***	0.0104***	0.0120***	0.00893***	0.00932***	0.0114***	0.00791***
	(0.00115)	(0.00213)	(0.00131)	(0.00116)	(0.00208)	(0.00134)	(0.00118)	(0.00229)	(0.00132)
Liberal professions/white collar director/entrepreneur	0.00754***	-0.00130	0.0150***	0.00822***	0.00220	0.0135***	0.00727***	0.00176	0.0120***
	(0.00142)	(0.00272)	(0.00180)	(0.00143)	(0.00259)	(0.00184)	(0.00149)	(0.00260)	(0.00198)
Mother last occupation (housewife excluded):									
Blue collar	-0.00110	-0.00192	-0.000787	-0.00142	-0.00233	-0.00129	-0.00199*	-0.00390*	-0.00124
	(0.00121)	(0.00237)	(0.00128)	(0.00119)	(0.00230)	(0.00126)	(0.00114)	(0.00229)	(0.00127)
Self-employed/small businessman	0.00275*	-0.00890***	0.0100***	0.00322**	-0.00594**	0.00884***	0.00397**	-0.00350	0.00846***
	(0.00164)	(0.00308)	(0.00185)	(0.00158)	(0.00296)	(0.00179)	(0.00156)	(0.00305)	(0.00181)
White collar	0.000641	-0.00747***	0.00668***	0.000706	-0.00634**	0.00581***	0.000520	-0.00591**	0.00521***
	(0.00130)	(0.00263)	(0.00140)	(0.00129)	(0.00259)	(0.00136)	(0.00126)	(0.00263)	(0.00138)
Liberal professions/white collar director/entrepreneur	-0.00758***	-0.0172***	0.00119	-0.00731***	-0.0155***	-0.000504	-0.00778***	-0.0141***	-0.00252
	(0.00213)	(0.00373)	(0.00245)	(0.00208)	(0.00369)	(0.00237)	(0.00204)	(0.00418)	(0.00226)
Municipal Variables									
Female mayor	0.00888	0.0104	0.00703	0.00843	0.00964	0.00722	0.00692	0.00717	0.00704
	(0.00589)	(0.00903)	(0.00469)	(0.00574)	(0.00843)	(0.00482)	(0.00533)	(0.00789)	(0.00460)
Share female councillors	0.0121	0.00555	0.0199*	0.0127	0.0103	0.0186*	0.0120	0.0141	0.0165
	(0.0112)	(0.0180)	(0.0113)	(0.0111)	(0.0175)	(0.0113)	(0.0113)	(0.0180)	(0.0114)
Fertility rate	5.36e-05	8.37e-05	4.64e-05	7.57e-05	0.000101	8.70e-05	1.81e-05	0.000123	3.46e-05
	(0.000101)	(0.000198)	(0.000111)	(9.91e-05)	(0.000193)	(0.000109)	(9.81e-05)	(0.000199)	(0.000109)
Share of religious marriages	-0.00686	0.00319	-0.0111**	-0.00860*	0.000272	-0.0123**	-0.00470	0.00323	-0.00942*
	(0.00517)	(0.0102)	(0.00549)	(0.00507)	(0.00992)	(0.00542)	(0.00502)	(0.0101)	(0.00545)
Supply of STEM courses	-0.00422***	-0.00644***	-0.00259	-0.00408**	-0.00621***	-0.00235	-0.00360*	-0.00596***	-0.00211
	(0.00162)	(0.00142)	(0.00227)	(0.00169)	(0.00136)	(0.00230)	(0.00195)	(0.00159)	(0.00236)
Supply of university courses	0.00102**	0.00160***	0.000644	0.000978**	0.00155***	0.000541	0.000792	0.00146***	0.000417
	(0.000448)	(0.000467)	(0.000666)	(0.000465)	(0.000431)	(0.000672)	(0.000523)	(0.000463)	(0.000681)
R-squared	0.143	0.142	0.086	0.203	0.211	0.135	0.244	0.270	0.181

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Notes: The sample consists of college graduates who were 18 between 2003 and 2010 who graduated between 2010 and 2015. The dependent variable is the maths intensity index of the course of study. Each regression includes the survey year, year of graduation and municipality of residence fixed effects.