Department of Economics and Statistics "Cognetti de Martiis"
Campus Luigi Einaudi, Lungo Dora Siena 100/A, 10153 Torino (Italy)
www.est.unito.it

Dipartimento Economia e Statistica Cognetti de Martiis

Working Paper Series

1/23

GENDER GAP IN HIGH SCHOOL CHOICES: DO ACHIEVEMENTS AND PEERS PLAY A ROLE?

DALIT CONTINI, MARIA LAURA DI TOMMASO, ANNA MACCAGNAN and SILVIA MENDOLIA

UNIVERSITÀ DEGLI STUDI DI TORINO

The Department of Economics and Statistics "Cognetti de Martiis" publishes research papers authored by members and guests of the Department and of its research centers. ISSN: 2039-4004

Gender gap in high school choices: do achievements and peers play a role?*

Dalit Contini†, Maria Laura Di Tommaso‡, Anna Maccagnan§, Silvia Mendolia**

Abstract

This paper focuses on the gendered choice of high school in the Italian context, where students select the type and subject-focus of their schools when they are around 13 years old (at a younger age than in most of the rest of the developed world). We explore the role of own ability and past performance in mathematics and Italian, the role of own position in the class ranking and the overall performance of girls and boys in the class. The main results show that individual ability by subject is an important determinant of high school choice and these effects vary with gender. Further, relative performance by subject also matter in educational choices, and the effects are similar for boys and girls. Lastly, we show that the gender gap in the choice of high school does not change once we have taken into consideration all these various mechanisms, and therefore the differences in choices are mostly unexplained, and likely to be related to cultural and social stereotypes, gender norms, and different perceptions of own abilities. These results call for policies that address stereotypes and prejudices.

JEL codes: I21, I24, J16.

Keywords: Gender gap, school choice, peer effects, abilities, limited dependent variable models

* Corresponding author: Silvia Mendolia (silvia.mendolia@unito.it)
† Dept. of Economics and Statistics Cognetti de Martiis, University of Torino.
‡ Dept. of Economics and Statistics Cognetti de Martiis, University of Torino and Collegio Carlo Alberto.
§ Dept. of Economics and Statistics Cognetti de Martiis, University of Torino.

** Corresponding author. Dept of Economics, Social Studies, Applied Mathematics and Statistics. Silvia.mendolia@unito.it

1. Introduction

Gender differences in educational choices (and, in particular, differences in the propensity to choose STEM fields of study) have attracted the attention of economists and social scientists for several years, as they have important long-term effects on career paths of men and women. Narrowing the gender gap in STEM education would lead to increased number of jobs and increased GDP over the long-term (EIGE, 2018). Variation of gender segregation exists across sub-fields in STEM, with ICT and engineering showing the largest gaps (21% of ICT graduates and 28% of Engineering graduates are female). Similarly, men are significantly under-represented in Education, Health and Welfare fields (18% of Education graduates and 24% of Health and Welfare graduates are male). This educational segregation is mirrored in a horizontal segregation in the labour market: at the EU level (EU-28 countries), women represent 75% of the workforce in Education, Health and Welfare occupations and 15% of the workforce in STEM related occupations (EIGE, 2018).

Existing literature has analysed possible determinants of these gaps, including gender gaps in achievements by subject, individual relative performance in various subjects, gender composition in the classroom, comparison effects, and gender stereotypes.

Gender gaps in achievements by subject have been investigated in the recent literature. In particular, several studies have shown that gender differences in previous mathematics performance do not drive gendered patterns in educational choices (see for example Card and Payne, 2021; Friedman-Sokuler and Justman, 2016; Kahn and Ginther, 2017; Rapoport and Thibout, 2018; Justman and Mendez, 2018, among many others), but rather these differences depend on how boys and girls interpret signals regarding their ability in the various subjects.

An interesting strand of recent studies have shown that relative performance in various subjects matters more than absolute performance. Girls generally have a comparative advantage in literacy-related subjects, and this drives their educational choices away from maths-intensive fields. Further, the gender gap in educational choices cannot be explained by differences in math ability across genders, but is significantly reduced when one controls for the gender gap in the individual differences between math and literacy performance (Breda and Napp, 2019; Mostafa, 2019; Stoet and Geary, 2018). The vast majority of this literature analyses performance during or at the end of high school and educational choices in university.

The impact of gender compositions on various students' outcomes has received increasing attention from economists in recent years. Several studies have analysed the impact of being placed in a classroom with a higher proportion of female or male peers, but results are unclear, and vary with the country's institutional settings, education and labour market policies, cultural values etc. For example, Hoxby (2000) and Lavy and Schlosser (2011) show that a higher proportion of female peers has beneficial effects for both genders, while, for example, Black et al (2013) show that increased percentages of females in the classroom is detrimental for boys' long run educational performance. Similarly, Schneeweis and Zweimüller (2012); Booth et al. (2014); Brenøe (2017); Schone et al., 2019 (among others) use random variation of gender composition across cohorts to identify the impact of gender composition of peers on students' performance and choices but the existing results do not show a clear direction of the effects. Recently, Anelli and Peri (2019) use a dataset of 30,000 Italian students to show that gender composition of high school peers has an effect on men but not on women, using random assignment of Italian pupils to first grade in high school. Interestingly, Brenøe and Zolits (2020) presents new findings, which contribute to the existing mixed evidence, showing that having a large proportion of female peers in high school reduces the chances of enrolling in a STEM program at university for female students.

Comparison effects are also important in affecting students' choices, but the evidence on this is very limited. One interesting study in this field is Cools et al (2019), who analyse the impact of exposure to female and male "high achievers" in high school on long run outcomes and show that greater exposure to "high-achieving" boys reduces girls' self-confidence and the likelihood to enrol in an undergraduate degree. However, in this study, "high achievers" are defined using parental education, rather than individual ability *per se*. Similar results are also presented in Mouganie and Wang (2020) and Feld and Zolitz (2018). Pagani and Pica (2021) use Italian data to show that students who have been exposed to a high share of academically gifted peers of the same gender in primary school are more likely to choose a scientific high school track, while the effect of peers of the other gender is the opposite.

An additional important factor which has been considered by the literature is the role of parents and teachers and the effect of gender related stereotypes and, more generally, cultural models (Nollenberger et al., 2016). For example, Carlana (2019) shows that the gender gap in maths performance dramatically increases when students have teachers with strong gender stereotypes (measured using the Gender-Science Implicit Association Test) and that g+irls are less likely to choose more demanding tracks and fields of study when their maths teachers have

strong stereotypes. Similarly, Carlana and Corno (2021) use an experimental context to demonstrate that children's choices are influenced by parental perceptions and gender stereotypes, as well as by the chances of interacting with peers of the same gender.

This literature is related to the broader field of study looking at men's and women's performance in situations with different levels of competitiveness, generally showing that women tend to balance their choices and reduce risks when possible (Delaney and Devereux, 2021), and are less likely to be willing to engage in highly competitive situations (see for example Niederele and Vesterlund, 2010; Niederle et al., 2013; Buser et a., 2014; Buser et al., 2017; Landaud et al., 2018, among others) and that men increase their self-confidence when they are competing with other men (Bengtsson et al, 2005).

This study analyses educational choices of Italian girls and boys at age 13, when the children leave middle school and choose high school. In the Italian education system, there is a substantial variation among high schools, in terms of subjects studied, level of difficulty, higher education pathways, and academic prestige (see section 2 for details). Further, students are free to choose the type of high school they prefer without any binding teacher's recommendation and there are huge differentials in school choices across social strata and gender. We focus on both the type of school i.e. more/less academic - and subject studied i.e. STEM vs Humanities or Other focuses. In particular, we analyse the determinants of these choices, including absolute and relative past performance in Italian and Maths, as well as peers' performance, and we consider whether any of these mechanisms are useful in explaining the gender gaps in educational choices.

We expand the existing literature in the area on gender gaps in subject choice at university, by providing new evidence on the existence and possible consequences of gender gap at a (relatively) young age. This is a very important contribution, as the majority of existing studies focus on choices of university studies at the end of high school. However, the roots of the gender gap in educational choices are likely to be found at a young age (Fryer and Levit, 2010), and policies are more likely to be effective if they target children at an age where skills and preferences are still relatively malleable. An analysis of the gender gap in middle/high school is especially relevant for kids who do not go to university and enter the labour market straight after high school. In particular, in Italy only 28% of the population 25-34 has a tertiary education attainment (OECD 2021).

Further, educational and social sciences literature has shown that young women from low socio-economic families are particularly vulnerable to elements driving students away from STEM studies and are generally less likely than their advantaged female peers to select these subjects (see for example Archer et al., 2012; Codiroli Mcmaster, 2017). We specifically investigate the gender gap in educational choices for children choosing academic, as well as non-academic tracks, and therefore this analysis can be useful to understand some of the reasons behind the gender gaps and segregation in low and middle-range occupations. This analysis is particularly relevant for children coming from low socio-economic families, who are less likely to pursue university studies.

An additional important element in our study is the analysis of the role of individuals' absolute and relative rank in affecting students' high school choices. This is a relatively unexplored area. Existing studies have looked at the effect of individuals' class rank on important long term educational and labour market outcomes, but have not focused on different impact of ranking across genders (see for example Murphy and Weinhardt, 2020; Denning et al., 2018, among others). Delaney and Devereux (2021) analyse the importance of individuals' ranks in English and Mathematics in explaining major college choices, and show that the tendency for girls to be lower ranked in maths contribute to explain part of the STEM gender gap. Goulas et al. (2022) find that girls assigned to a high school class where they have a comparative advantage respect to their peers are more likely to choose a STEM subject in tertiary education

We use a new administrative dataset, which includes observations of Italian children in middle school, when they choose the track for secondary education, and covariates for family's and school's characteristics. This dataset includes extensive information on the children's academic performance in various subjects, including teachers' evaluation as well as standardized test scores.

Our main findings suggest that the gender gap in the choice of high school is not explained by past performance differences. We show that girls are less likely to select STEM studies, and these gaps are particularly relevant for middle range school types. We also know that boys are less likely to select Humanities oriented schools, even when they have abilities in these areas. Math performance matters to choose a STEM High School both for boys and girls but the gender gap in the probability to choose a STEM oriented school does not change with past

performances. We have also analysed other possible explanations of these gender gaps, considering the role of past relative performance by subject, of comparative advantage and of peers' performance. The main conclusion of this paper is that none of these mechanisms reduce the gender gap in educational choices.

We have also decomposed the gender gap in educational choices into an explained and an unexplained component, utilising a procedure similar to the Oaxaca-Blinder decomposition of the gender wage gap. The results of the decomposition confirm the previous analysis and show that most of the gaps in educational choices are unexplained by observed characteristics. There must be other unobserved factors like gender stereotypes, social norms as well as parents and teachers bias that influence the choice of high school.

2. Institutional Context and Data

Formal education starts at age 6 in Italy and is organised in three stages. Students attend five years of primary school (until age 11), three years of middle school (age 11- age 14) and five years of high school (although compulsory education ends at age 16).

The main criterion to allocate Italian students to primary and middle school is geographical proximity to their place of residence. Students should usually attend a school in their local catchment area, and within a school they are (almost randomly) assigned to a specific class (group of students with a specific set of teachers), which remains the same for all the primary (or middle/high) school years. Therefore, this process should not allow for specific requests in terms of teachers or classmates. Practically, some exceptions are allowed, and schools may consider family requests (especially for family's specific circumstances, such as proximity to parents' place of work or grandparents' residence) on a case-by case basis.

Middle school curriculum is the same in all Italian schools and classes are usually formed to be heterogeneous in terms of gender and prior achievements. No streaming by ability occurs in the Italian school system and students do not select elective subjects.

At the end of middle schools, students sit a national exam and transition to high school. Middle school teachers usually provide families with some broad suggestions regarding high school choices, but high schools cannot select students based on prior academic ability. Further, students are free to apply to any high school they prefer, without necessarily considering teachers' suggestions.

High schools in Italy can be classified in two ways. First, in terms of the main subject studied (Horizontal classification): STEM, Humanities, and other content, which is largely economics, but can vary in the case of vocational paths. Second, in terms of the level and prestige of the track (Vertical Classification): Traditional lyceum , Non traditional lyceum, Technical track, and Vocational track. Lyceums offer a generalist academic education, with students likely to continue their studies into tertiary education. Traditionally, lyceums were of two types, the scientific lyceum, focusing on STEM subjects, and the Classical lyceum, focusing on Humanities. New types of lyceums have been instituted, mainly the Linguistic lyceum, the Artistic lyceum and the Human science lyceum. This can also have a socio-economic focus. The technical track and the vocational track offer a more applied preparation.

Previous research (see for example Checchi & Flabbi, 2007, Contini & Scagni, 2013) has shown that family socio-economic background is important in the choice of the high school in Italy, and children from high socio-economic families are over-represented in Lyceums.

The horizontal and vertical high school classification cross is shown in Table 1

Table 1 - High school classification

	STEM	Other	Humanities
Traditional lyceum	Scientific lyceum		Classical lyceum
Non-traditional lyceum		Human sciences lyceum with socio-economic focus	Linguistic lyceum/ Artistic lyceum/ Human sciences lyceum
Technical track	Technical paths e.g. Informatics, Chemistry, Electronics	Technical paths e.g. accounting, marketing	
Vocational track	Vocational paths e.g. Agricultural or Mechanical operator	Vocational paths e.g. Commercial operator, Catering school, Hotel management school	

This study uses a longitudinal dataset linking together data from the Italian National Register of Students (Anagrafe Nazionale Studenti), and the National Institute for the Evaluation of the Educational System (INVALSI).

The linking allows us to have a complete picture of students' scholastic achievements prior to choosing high school – i.e. test scores and teachers' grades - and careers, together with

variables capturing their socioeconomic background[6]. In terms of students' careers, we have information on dropouts, grade repetitions, school and class changes, as well as high school choice – which the main focus of this research. Thanks to the presence of class and school identifiers, we can also link information on classmates' and schoolmates' scholastic attainment and background.

Data are available for secondary students enrolled in school years 2013/14 to 2016/17 in three Italian regions: Piedmont, Lombardy and Veneto. Given the purpose of this analysis, we work on the cohort of students enrolled in first year of middle school in 2013/14 (year 6 – prima media). In school year 2016/17 these students start their first year of high school (year 9 – prima superiore), when tracking begins.

While the complete sample is made up of 173,684 students, we can only observe high school choices for 168,445 students. Interestingly, this dataset includes information on standardized test scores, as well as teachers' grades in Italian and Mathematics.

Table 2 and 3 present descriptive statistics of the estimation sample and shows that boys are over-represented in schools focusing on STEM subjects, as well as in Lyceums. Further, they are clearly more likely to attend both Traditional and Technical STEM schools. Around 30% (20%) of boys (girls) in the estimation sample attend Traditional STEM schools, and the gender gap is even more noticeable in Technical STEM schools, which are chosen by over a third of the boys in the estimation sample, and by only 7% of girls.

Table 4 presents gender gaps in academic performance, measured by test scores and teachers' grades. Interestingly, girls overperform with respect to boys in both Italian and Mathematics, when performance is measured by teachers' grades, but their performance in standardized Math test scores is significantly lower than the boys' one.

We exploit the richness of our dataset and include a wide set of control variables, including individual (gender, grades, test scores, migration status), family (parental education and an index of socio-economic status) and middle school (Proportion of students with parents with university degree; proportion of migrants in class; average class test scores in maths and Italian) characteristics.

[6] Using this joint data set respect to use only INVALSI data allows us to have teachers'grades at the end of the school year and the type of high school students choose at age 14. Respect to using only ANS (anagrafe nazionale studenti) we are also able to use standardised test scores in grade V and grade VIII and the socio economic status of each student.

Table 2 –Horizontal and vertical classification of high schools

Horizontal classification (subject)

	Males (%)	Females (%)
STEM	*69.87*	*28.59*
Other	*19.56*	*31.55*
Humanities	*10.57*	*39.87*

Vertical classification (type of school)

	Males (%)	Females (%)
Traditional lyceum	33.45	26.65
Non-traditional lyceum	9.20	37.90
Technical track	47.83	24.76
Vocational track	9.52	10.69

Table 3 – Educational choices

	Males (%)	Females (%)
Traditional STEM	30.41	20.33
Traditional Humanities	3.05	6.32
Non-trad/Tech other (econ)	13.24	22.33
Non-traditional Humanities	7.53	33.54
Technical STEM	36.27	6.78
Vocational STEM	3.19	1.48
Vocational Other	6.32	9.21
Total	100	100

Table 4 – Grades and Test scores by gender and subject

	Males	Females
Test score in Italian (standardized)	0.153	0.384
Test score in mathematics (standardized)	0.404	0.154
Grade in Italian (0-10)	7.185	7.592
Grade in mathematics (0-10)	7.367	7.495

3. Econometric methods

We estimate a multinomial logit model with M discrete choices of high school:

$$Pr(Y_i = j|x_i) = \frac{exp\ (x_i\beta_j)}{1+\sum_{m=1}^{M-1} exp(x_i\beta_m)} \quad (1)$$

where the vector X includes gender, individual grades and test scores as well as individual, parents' and school's characteristics. Further, the model includes interactions between the gender dummy and test scores and grades.

In the first part of the study, we analyse whether the gender gap is explained by test scores, grades, and socio-economic characteristics. For this reason, we begin by estimating a simple multinomial logit where only the gender dummy is included. We then compare the size and significance of the coefficient of the gender dummy in several versions of the model, where more covariates are progressively included.

Next, we study whether grades, test scores, rankings and other mechanisms affect the educational choices of girls and boys differently, by estimating a multinomial logit including all these covariates.

The main focus of the analysis is on understanding if and how the gender gap in educational choices can be explained by past performance. Therefore, we investigate whether this gap is attenuated when several transmission mechanisms (grades, test scores, ranking, etc) are included in the model.

An important assumption in this context is that test scores/grades or other mechanisms are uncorrelated with the error term, conditional on all the controls. This assumption can be violated if students can sort themselves into classrooms based on their past performance or if there are unobserved variables (such as, for example, individual effort, self-confidence, etc.) that affect past achievements and educational choices.

First of all, the Italian institutional context is based on students' attendance to a middle school that is close to their area of residence, and schools form classroom groups with a specific attention to create heterogenous groups, with an equal representation of students from different gender and levels of ability. Therefore, it is hard for students and parents to select schools on the basis of academic performance.

Further, the main focus of our analysis is on the gender differences in the impact of the various variables (test scores, grades, ranking etc), as we are mostly interested in analysing whether any of these mechanisms help explain gender differences in educational choices (rather than on the causal effect of test scores or grades on high school choices).

However, we follow the literature in the field (see for example Rapoport and Thibout, 2018, among others) and perform a sensitivity test using test scores obtained at the end of primary school as an instrument for middle school performance. The identifying assumption here is that children do not choose their levels of effort and performance in Italian and Maths in primary school anticipating future choices in high school (Rapoport and Thibout, 2018) and that, once we control for all individual, school and family observable characteristics, primary school performance per se does not have a direct effect on educational choices made later in life. Further, test scores at the end of primary school are not important in students' records and do not affect students' chances to attend a particular middle (or high) school, therefore we expect students and parents to be less likely to invest a significant time and effort to affect results of these tests. Lastly, Invalsi test scores are made up of several multiple-choice questions and, in the majority of cases, of questions where students have to select the right answers. Even when students must answer open-ended questions, teachers receive very clear marking guides that should limit the risk of bias in grading.

It is also possible that middle school teachers have unobservable characteristics that affect students' choices with gender-specific effects, because of specific teaching practices or stereotypes (Carlana, 2019). We try to mitigate these effects by including in the model the class average and standard deviation of Italian and Mathematics test scores.

Finally, we consider the importance of school's unobserved characteristics by estimating a version of the model including school fixed effects. Given the computational difficulties of estimating a multinomial logit model with 7 choices of high school including school fixed effects (see Appendix), we estimate results with fixed effects for a subsample of students from one city only (available on request) and observe that these are very similar to results where school fixed effects are not included, but we include several school characteristics (such as average academic performance and socio-economic background of the children in the school). Therefore, we present results without school fixed effect throughout the paper.

In addition, we decompose the gender gap in the probabilities of choosing each specific type of high school using a Blinder-Oaxaca like decomposition that can be applied to categorical

variables. Following Fairlie (2005, eq 2.2), the explained component is computed as the difference between the observed proportion for one of the two groups (say, boys) and the counterfactual predicted proportions that would be experienced by girls if they displayed the same behavior of boys (obtained by multiplying the girls' characteristics by boys' estimated coefficients of the multinomial regression). The unexplained component is derived as the difference between the total gap and the explained component.

4. Results

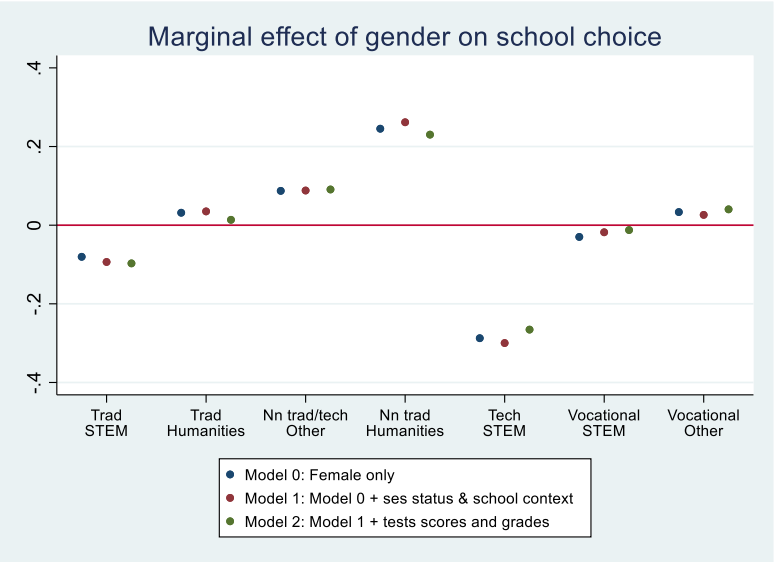
The first purpose of our analysis is to investigate whether the gender gap in educational choices can be explained by past achievements and other observable characteristics, such as indicators of individual performance and peers' characteristics.

For this reason, we begin by estimating a multinomial logit model of the probability of selecting each type of high school, depending on individual gender, and other observable variables. In the first model, we only include an indicator of gender. Then, we gradually add more covariates, such as an indicator of individual socio-economic status, parental education and migration background. We also add some variables related to the school's context, such as the proportion students with parents with university degree, proportion of immigrant students, the class' average and standard deviation of Italian and Maths test scores, and the proportion of girls in the classroom.

Last, in the most complete version of the model, we add individual grades and test scores in Italian and Maths, in order to understand if, and how, the estimated coefficients of the gender binary variable change when we include these measures of past achievements.

We visualise below in Figure 1 some interesting findings, showing that the gender gap in educational choices is very similar throughout the various models. Figure 1 shows that girls are significantly less likely to select STEM schools, including the more academically demanding ones (Traditional STEM) and, at an even higher degree, the Technical STEM paths. On the other hand, they are disproportionately more likely to select Humanities-focused schools (Traditional or non-Traditional Lyceum with Humanities focus) and these results do not vary when test scores, grades and other observable characteristics are included in the model.

Figure 1 – Marginal effects of gender on school choice



These results show that past achievements in Italian and Maths and other observable characteristics ***do not*** explain the gender gap in educational choices.

Therefore, we next analyse the impact of several transmission mechanisms on the educational choices of girls and boys. We are interested in understanding ***how*** these mechanisms change educational choices and ***if*** the impact varies across gender.

In particular, we investigate the role of:

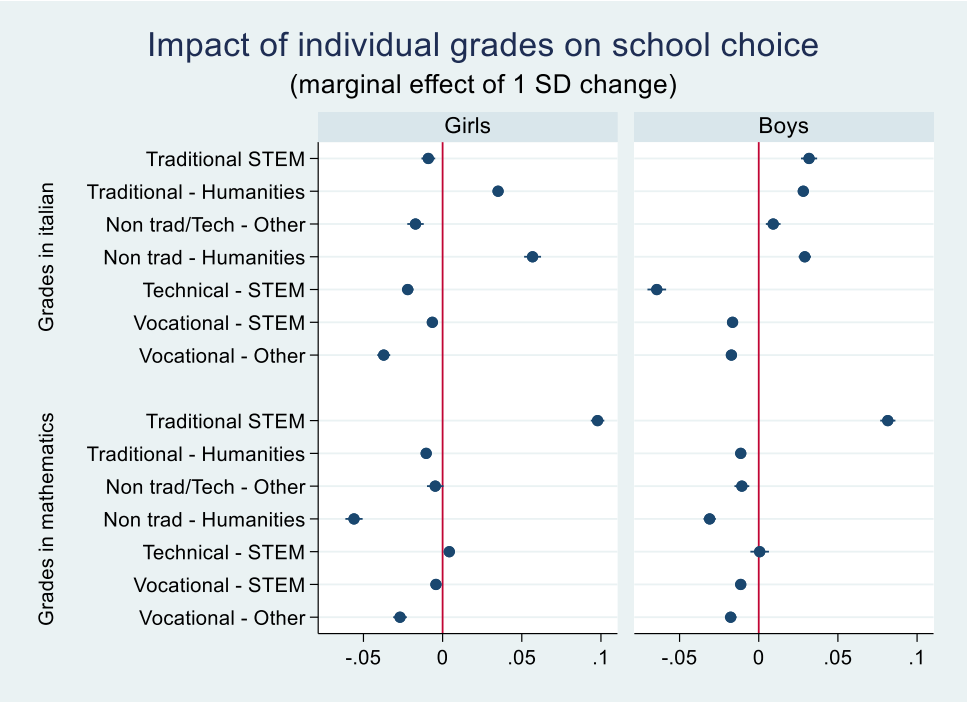
- Individual achievements by subjects
- Individual ranking by subjects
- Individual comparative advantages by subjects
- Gender roles and peers' performance

Figure 2 presents the marginal effects of 1 standard deviation change in individual grades in Italian and Maths on educational choices. The complete tables of estimates and marginal effects are presented in the Appendix (see Tables A1 to A3). Figure 2 shows that increasing Italian grades, students are more likely to choose Traditional Lyceums (rather than other schools). However, boys are more likely to select Traditional STEM, while girls are more attracted by Humanities-focused schools. Therefore, Italian grades have opposite effects on the choices of boys and girls, especially with respect to STEM-oriented schools.

On the other hand, increasing Math grades drives both boys and girls towards Traditional STEM studies. The sizes of the effects are broadly similar across gender, and we note that

increasing Math grades has a negative effect on the probability of selecting all high schools different from the Traditional STEM path. Clearly, these results show that the "Vertical" dimension of school choice is as important as the "Horizontal" one and students from the highest part of the grade distribution are more likely to select more academically challenging schools.

Figure 2 – Impact of grades on educational choices



Next, we calculate the predicted probabilities of boys and girls to select various types of high schools, depending on their grades in Italian and Maths. This is useful as it allows understanding the predicted choices of students in various parts of the grade distribution for both Italian and Maths and analysing how these choices may vary by gender.

Results are presented in Figure 3a and 3b, where we show predicted choices at various levels of performance in Italian and Mathematics.

Figure 3a – Predicted probabilities of educational choices by Italian and Maths test scores (Females)

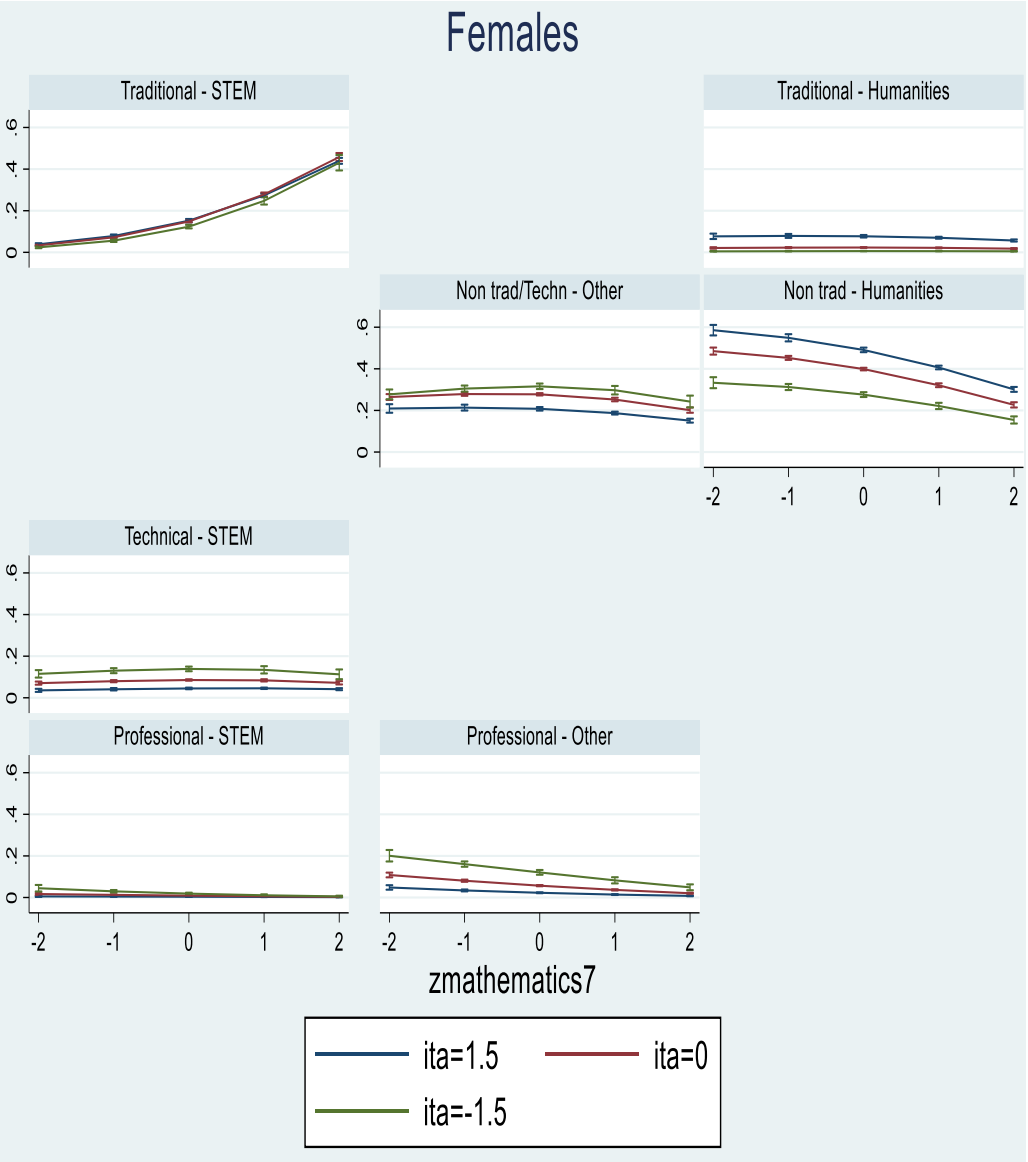
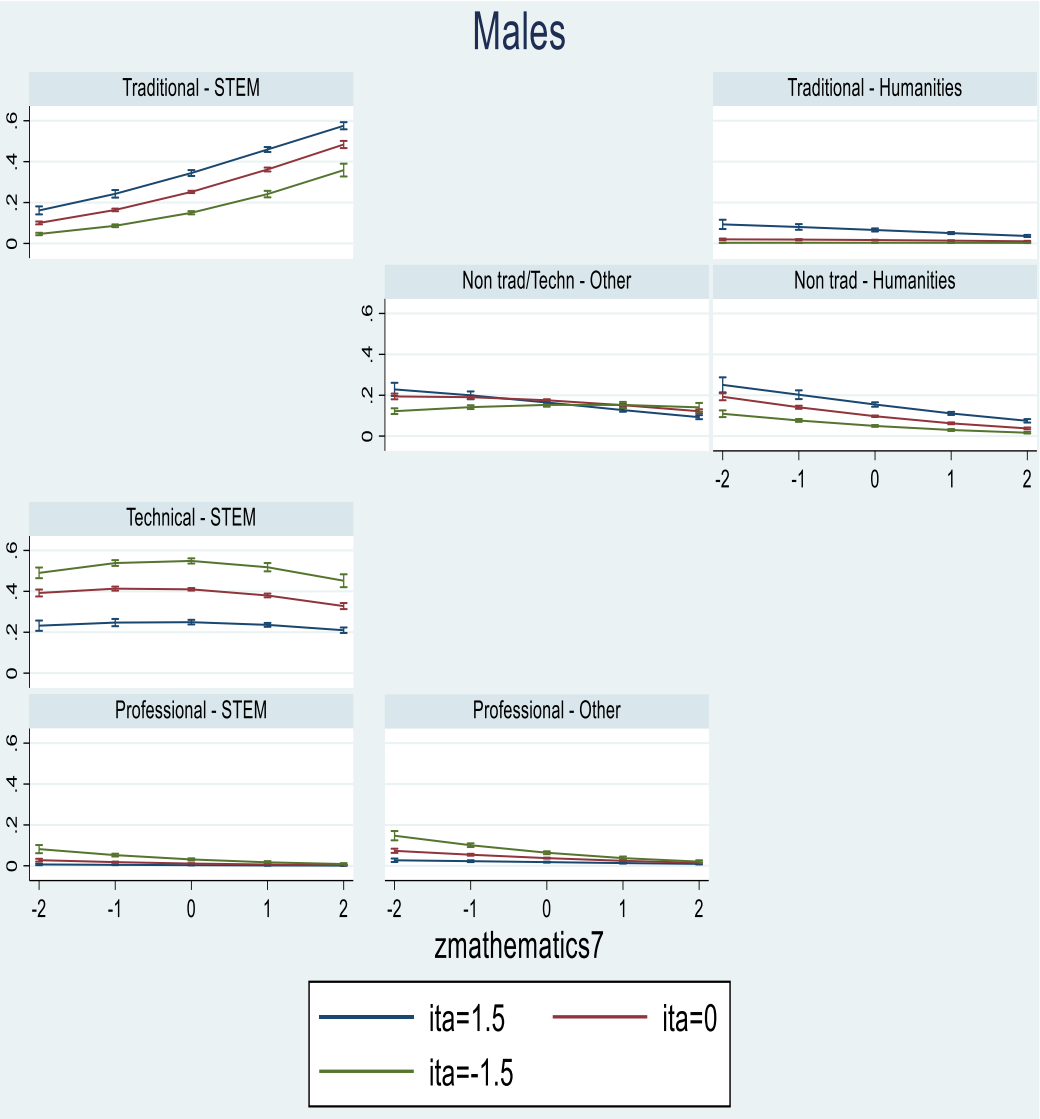


Figure 3b – Predicted probabilities of educational choices by Italian and Maths test scores (Males)



The complete table of probabilities is presented in Appendix Table A4. A boy in the top part of the ability distribution for both Maths and Italian has a predicted probability of selecting Traditional STEM studies of 63% and a probability of selecting Technical STEM of 16%. The same probabilities for a girl with similar achievements are 45% and 4%, while the predicted probabilities of selecting Humanities schools are substantially higher for girls than for boys (11% vs 7% for Traditional Humanities and 28% vs 6% for Non-Traditional Humanities).

These gender gaps are even more noticeable if we analyse students in the middle or the bottom part of the grade distribution. A boy with average performance in Maths and Italian test scores has a predicted probability to choose Traditional STEM (Technical STEM) equal to 24% (41%), while the same predicted probability for a girl with similar levels of ability is 14%

(9%). On the other hand, girls in this part of the test scores distribution are disproportionately more likely to choose Humanities-focused schools (the predicted probability of choosing Non-Traditional Humanities Lyceums is equal to 37% for girls and 9% for boys).

These results differ from previous findings in existing literature. For example, Rapoport and Thibout (2018) show that test scores reduce gender gaps in educational choices and that boys and girls with similar (high) grades in both subjects make relatively similar choices.

Overall, we observe that girls from the top of the maths distribution are the only ones who do not shy away from STEM studies (even if, when they excel in Italian too, they still choose humanities schools in high proportion). Further, these results show that boys choose STEM studies more than girls, even if their past performance in Maths is not excellent.

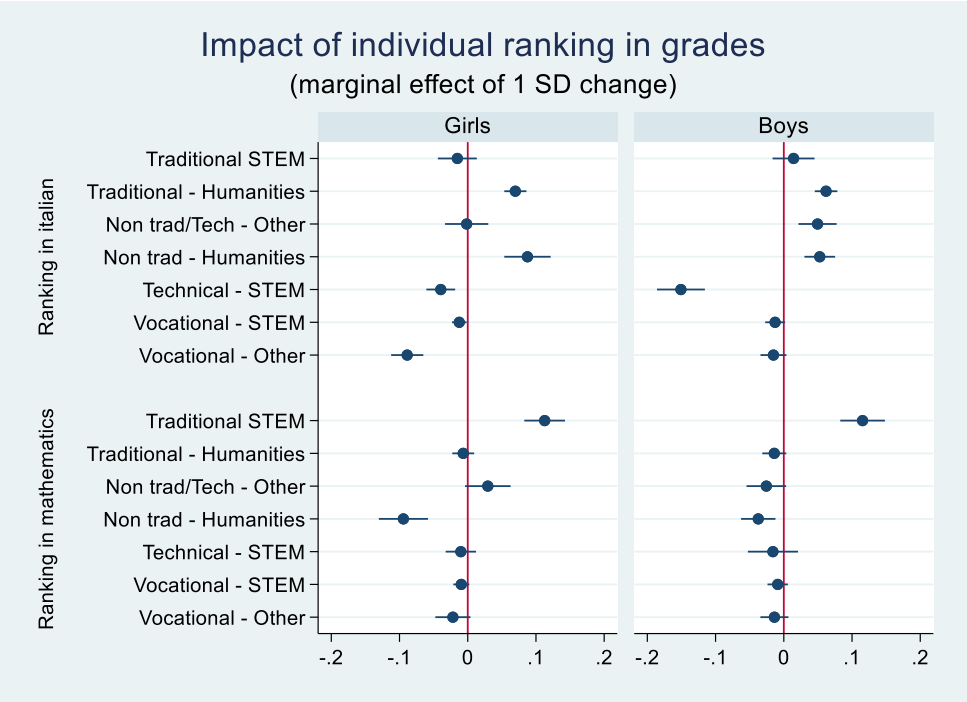
As explained in the Methodology section, we consider the possibility that test scores in year 8 are endogenous and use instrumental variable estimation in a model where test scores in year 5 (at the end of primary school, before students move to middle school) are used as an instrument for test scores in year 8. Results are reported in Appendix Tables A5 and A6 and are very similar to the ones presented above. Test scores in year 5 are highly correlated with test scores in year 8, and this is particularly true for results in the same subject, while the impact of year 5 test scores in Italian on year 8 test scores in Mathematics (and *viceversa*) is slightly lower (as expected). Results with instrumental variable estimation confirm previous findings. Italian grades have opposite effects for boys and girls, as increasing performance in Italian, boys are more likely to choose the Traditional STEM path while girls continue to prefer humanities-focused schools. On the other hand, increasing Maths test scores increases the probability f selecting STEM studies for both genders.

Next, we gradually include additional covariates to analyse the impact of other mechanisms in the educational choices of boys and girls and to see if they somehow explain the gender gap in choices. We begin by analysing the role of individual relative ability by subject, captured by the individual ranking by subject in the class.

Figure 4 presents the marginal effect of one standard deviation increase in individual ranking in Italian and Maths. Overall, the effects are as expected, and somehow similar to the effects of test scores and grades. Interestingly, rankings are significant in explaining educational choices, in addition to test scores and grades. This implies that relative performance is important (in addition to the absolute one). Increasing individual rankings in Italian increases chances to attend a Humanities-focused school and the effect is higher for girls than for boys.

On the other hand, increasing ranking in Maths drives students towards Traditional STEM studies and this is true for both genders. As already noted for absolute performance, relative performance in Mathematics increase students' confidence and seem to attract them towards more academically challenging schools. Complete results are reported in Tables A7, A8 and A9.

Figure 4 – Impact of rankings on educational choices



Lastly, we examine the role of two additional mechanisms in affecting the educational choices of boys and girls. First, we analyse the impact of students' comparative advantage in a particular subject. Previous literature has shown that differences in educational choices at university are mostly driven by girls' educational advantage in humanities-related fields, rather than by their disadvantage in STEM subjects (see for example Goulas et al., 2022 and Card and Payne, 2020).

For this reason, we estimate a model including the difference between Italian and Maths test scores (or grades). Complete results are reported Table A10, A11 and A12 and show that increasing the difference in test scores between Italian and Maths, students are less likely to choose the Traditional STEM path, and this is especially true for girls. However, these effects are overall quite small and gender differences are marginally noticeable.

Next, we investigate a final mechanism, related to peers' ability in the class. The idea here is to understand whether peers' performance by subject matters and whether the effect varies

depending on the peers' and individual gender. In particular, we introduce a new variable in the model, representing the proportion of girls with top math grades in the class. Results (see Appendix Table A13, A14 and A15) show that peers' ability matter for educational choices and increasing the proportion of high achievers in the class, students are less likely to choose the Traditional STEM study-path. However, these are relatively minor effects and we do not observe significant gender differences in the effects.

The analyses performed so far shows that the gender gaps in educational choices are significant, persist throughout the grade distribution, and are not attenuated when we take into account absolute and relative performance, as well as peers' ability and the school context.

Girls are systematically less likely to select STEM studies, unless they have a very strong comparative advantage in Mathematics. Further, they are especially less likely to select STEM studies when they are in the middle or low part of the test scores distribution.

The highest gaps are found for girls not choosing Technical-STEM education (in favour of Humanities-oriented schools) and the under-representation of girls in Technical STEM education is particularly worrying for its potential impact on future working life. Graduates from technical-STEM schools are more likely to find employment quickly after graduation and to have ongoing contracts (Fondazione Agnelli and CRISP, 2018) and, if they continue their studies, graduates from technical-STEM schools are more likely to continue STEM studies at university (Deloitte, 2021), with respect to those who attended other types of high schools.

Women continue to be significantly under-represented in STEM university studies and in particular in those fields with increasing demand in the labour market (IT, Engineering, etc) and therefore, these gaps in educational choices at a very young age are likely to have important effects on career paths and on segregation in the labour market. This is especially true for students with relatively poor performance and from lower socio-economic backgrounds, who are more likely to attend middle-range schools.

The previous findings have showed that gender gap in educational choices cannot be explained by past achievements, or the other mechanisms investigated. To further explore this main finding, we analyse whether girls' school choices would change if girls' test scores in maths were more similar to boys' test scores. We simulate an increase in girls' test scores for them to have the same average test scores of boys.

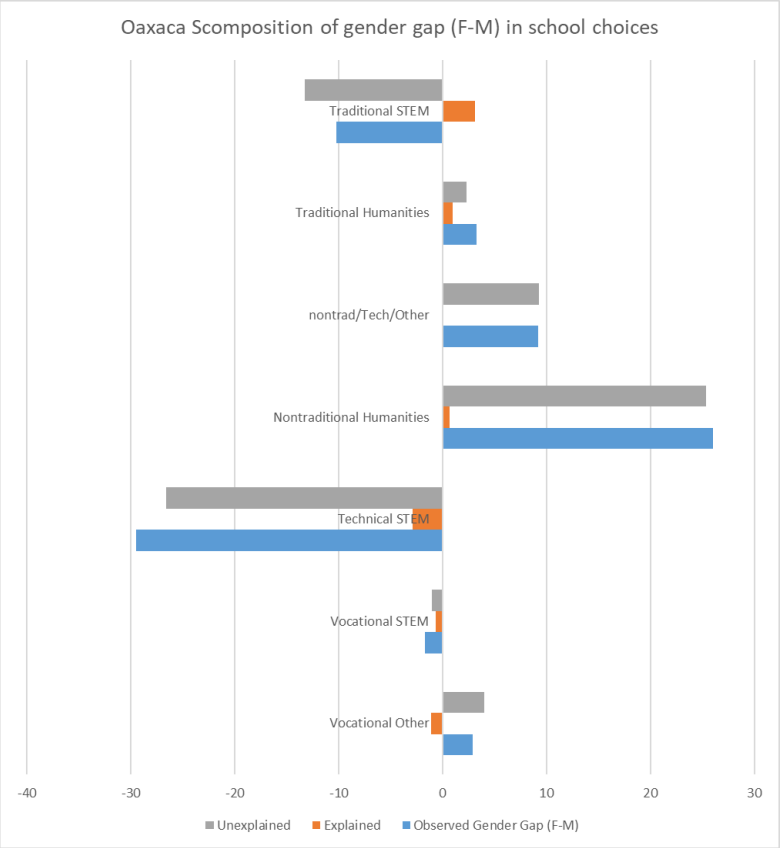
Results are presented in the Appendix (Table A16) and show that this change in girls' test scores distribution would have a sizeable effect on girls' probability to select Traditional STEM, which would increase from 20% to 23 % (reducing the current gender gap by around a quarter). However, there would not be any noticeable effect on the probability of selecting Technical STEM schools, where the gender gap would only be reduced by 2%. This confirms that past performance has a limited effect on educational choices and that there are many unobserved factors, such as gender biases and stereotypes, which are likely to play a major role in pupils' choices.

5. Oaxaca-Blinder decomposition of the gender gap in choice of high schools

We decompose the gender gap in the choice of high school using a procedure for categorical dependent variables (Fairlie, 2005) similar to the Oaxaca-Blinder decomposition. We use the coefficients of the model with IV (Table A6 in Appendix), and decompose the gap into an explained and an unexplained component. The explained component is obtained by multiplying their observed characteristics to the estimated coefficients resulting from the boys' model. These are the girls' expected probabilities of enrolling in each type of school if they behaved as boys (counterfactuals). Since the two groups share the same behaviour, while individual characteristics vary, we refer to this as the explained component of the gap. The explained component of the gender gap is then subtracted from the observed gap to obtain the unexplained component, i.e. the part of the gender gap that cannot be ascribed to the different composition of the two groups, but instead to their different behaviour.

Main results are shown in Figure 4 while the full results are included in Table A17 in the Appendix. The results suggest that the gender gap in educational choices is mostly unexplained by observed characteristics. In particular for traditional STEM, if girls' characteristics were remunerated with the same betas of boys, the gender gap would be in favour of girls.

Figure 4 – Oaxaca-Blinder decomposition of the gender gap in educational choices.



6. Conclusion and Discussion

We have analysed gender differences in the educational choices of Italian students, at age 12-13, between middle and high school. Our work complements the existing literature on gender gaps in education by shedding some light on the gendered patter of educational choices at a relatively young age, in an educational system where tracking happens earlier than in many other parts of the developed world.

We have also analysed the possible explanations of these gender gaps, considering the role of past (absolute and relative) performance by subject, as well as the effect of comparative advantage and peers' performance. The main take-home message from this analysis is that these mechanisms do not reduce the gender gap, which persists when we include these covariates in the model. The Oaxaca-Blinder decomposition of the gender gaps in educational choices confirms that the gaps are mostly unexplained by observed characteristics. Therefore,

the main sources of gender gaps in educational choices are likely to be found in unobservable factors, such as gender stereotypes, social norms and parents' and teachers' bias.

In particular, our work shows that test scores play an important role in affecting boys' and girls' choices, even if past performance has a different effect on boys and girls. Boys are more likely to select STEM studies even when their past performance in Mathematics is average, and continue to select STEM studies even when they achieve very high results in Italian. On the other hand, girls are systematically likely to select STEM-focused schools, unless they have a very strong comparative advantage in mathematics (i.e. very high math test scores, but not so high results in Italian).

The most striking result is the systematic difference in choices for pupils coming from the middle of the test scores distribution, where boys continue to choose STEM schools (even with lower levels of academic difficulty, i.e. the Technical STEM path) while girls continue to prefer humanities-focused schools. These gaps are particularly problematic, as they are likely to have important effects on labour market segregation and gender pay gaps.

These results are very relevant for all those countries where pupils' choices of high schools differ by gender and where the labour market is horizontally segregated by gender. In addition, our results are relevant for some countries like Belgium, Germany, Hungary and Austria, where the choice of high school happen at a younger age respect to other educational systems.

This paper shows that in order to reduce the gender gap in educational choices it is important to invest in policies that can deconstruct stereotypes and prejudices about women and STEM and men and Humanities. It also shows that children performances in numeracy and literacy do not play an important role in explaining the horizontal segregation in the choice of high schools.

References

Adkins, L.C. and Gade, M.N. (2012), "Monte Carlo Experiments Using Stata: A Primer with Examples", Terrell, D. and Millimet, D. (Ed.) 30th Anniversary Edition (Advances in Econometrics, Vol. 30), Emerald Group Publishing Limited, Bingley, pp. 429-477. https://doi.org/10.1108/S0731-9053(2012)0000030019

Anelli, M., Peri, G (2019). The effect of high school peers' gender on college major, college performance and income. The Economic Journal 129: 553-602.

Booth, A.L., Cardona, S.L. and Nolen, P. (2014). Do single-sex classes affect achievement? An Experiment in a Coeducational University, Discussion Papers No. 10221, CEPR.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2013). Under pressure? The effect of peers on outcomes of young adults. Journal of Labor Economics, 31 (1), 119-153.

Brenøe, A. (2017). 'High school gender composition and STEM choice in tertiary education', mimeo, Copenhagen University.

Brenøe and Zolits (2020). Exposure to More Female Peers Widens the Gender Gap in STEM Participation. Journal of Labor Economics, forthcoming.

Buser, T., Niederle, M., Oosterbeek, H (2014) Gender, Competitiveness and career choices. Quarterly Journal of Economics 2014: 1409-1447.

Carlana, M. (2019) Implicit Stereotypes: Evidence from Teachers' gender Bias. Quarterly Journal of Economics 2019: 1163-1224.

Cools, A., Fernandez, R., Patacchini, E. (2019) Girls, Boys, and High Achievers. IZA DP 12314.

Delaney, J. Devereux, P. (2021) High School Rank in Math and English and the Gender Gap in STEM. Labour Economics, forthcoming.

Delaney, J. Devereux, P. (2021) Gender Differences in College Applications: Aspiration and Risk Management. Economics of Education Review.
https://doi.org/10.1016/j.econedurev.2020.102077

Denning, J., Murphy, R., Weinhardt, F. (2018) Class Rank and Long-Run Outcomes. IZA DP 11808.

Fairlie R.W. (2005) An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models, Journal of Economic and Social Measurement, 30(4), 305-316

Feld, J., and Zolitz, U. (2018). Peers from Venus and Mars: Higher-Achieving Men Foster Gender Gaps in Major Choice and Labor Market Outcomes. Working Paper.

Goulas et al. JHR forthcoming 2022

Hoxby, C. M. (2000b). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867.

Iaria, A. and D'Haultfoeuille, X. (2016). A convenient method for the estimation of the multinomial logit model with fixed effects. Economics Letters, 141, 77-79. https://doi.org/10.1016/j.econlet.2016.02.002

Lavy, V., and Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. American Economic Journal: Applied Economics, 3 (2), 1-33.

Mouganie, P., and Wang, Y. (2017). High Performing Peers and Female STEM Choices in School. Working Paper.

Murphy, R., Weinhardt, F (2020). Top of the Class: The Importance of Ordinal Rank. The Review of Economic Studies, 87(6): 2777-2826.

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.

Schneeweis, N. and Zweimüller, M. (2012). 'Girls, girls, girls: gender composition and female school choice', Economics of Education Review, vol. 31(4), pp. 482–500.

Appendix

Empirical challenges in estimating a multinomial logit model with school fixed effects

We investigated the possibility to estimate a multinomial logit model with Fixed Effects to deal with unobserved heterogeneity at the level of the class of origin in year 8.

However, this has proved to be computationally extremely challenging, as the time and memory required to do so increase exponentially with the number of alternative school choices and the size of the class in year 8. As explained by D'Haultfoeuille and Iaria (2016), the estimation of such model is carried out through a Conditional Maximum Likelihood estimator and requires the listing of all possible permutations of the observed sequence of students' school choices within class of origin. With seven possible high school choices, and an average class size of … students, with a maximum size of 32 students, this has resulted in STATA running out of memory when estimating our full model.

Two different alternative strategies have therefore been explored.

First, we have estimated our multinomial logit models with fixed effects on random samples of the total numbers of permutations. As suggested by D'Haultfoeuille and Iaria (2016), in fact, permutation reduction is an effective way to cut estimation times and to run estimates that would otherwise not be feasible, although this may be costly from an efficiency point of view. To test the consistency of our results, the estimate has been repeated on several different random samples of permutations.

A second strategy consists in cutting the number of permutations by reducing either the number of alternative school choices, or the size of the class of origin. As for the former, we have run our multinomial logit models with fixed effects separately on the three vertical school categories (Lyceums, Technical and Vocational), and the three horizontal school categories (STEM, Humanities, and Other)[7]. These results have been compared to the results of a multinomial logit model without fixed effects.

As for the latter, i.e. the class size reduction, we have run our multinomial logit models with fixed effects on random samples of 8 students from the same class of origins. In particular, we have used the postfile command in STATA, that allows to carry out a Monte Carlo Simulation using ten repetitions each time, and then presented the simulated estimated parameters (Adkins and Gade, 2012).

[7] For computational reasons, this option has been combined with the permutation reduction.

Table A1 – Impact of grades and test scores on educational choices (Coefficients)

	Traditional Humanities	Non Trad./ Tech. Other	Non Trad Humanities	Technical STEM	Vocational STEM	Vocational Other
Female	0.856***	0.999***	1.872***	-1.069***	0.293***	0.917***
	(0.056)	(0.027)	(0.029)	(0.029)	(0.075)	(0.041)
Grade in Italian	0.715***	-0.228***	0.125***	-0.510***	-1.032***	-0.703***
	(0.039)	(0.025)	(0.028)	(0.020)	(0.058)	(0.038)
Female* Grade in	0.085	0.045	0.026	0.094*	0.425***	0.117*
Italian	(0.050)	(0.033)	(0.034)	(0.037)	(0.085)	(0.050)
Grade in	-0.563***	-0.504***	-0.757***	-0.427***	-0.931***	-0.806***
mathematics	(0.039)	(0.024)	(0.029)	(0.019)	(0.057)	(0.039)
Female* Grade in	-0.136**	-0.203***	-0.057	-0.197***	-0.097	-0.241***
Mathematics	(0.049)	(0.033)	(0.035)	(0.036)	(0.087)	(0.051)
Test scores in	0.631***	-0.174***	0.190***	-0.395***	-0.790***	-0.557***
Italian	(0.042)	(0.028)	(0.033)	(0.023)	(0.049)	(0.039)
Female* Test score	0.043	-0.024	0.004	0.121**	0.232**	-0.027
In Italian	(0.052)	(0.037)	(0.039)	(0.041)	(0.081)	(0.051)
Test scores in	-0.476***	-0.541***	-0.653***	-0.220***	-0.609***	-0.695***
Mathematics	(0.038)	(0.023)	(0.027)	(0.018)	(0.042)	(0.032)
Female* Test score	-0.048	0.013	0.056	-0.070*	0.040	-0.046
In Mathematics	(0.047)	(0.031)	(0.032)	(0.034)	(0.069)	(0.043)
Escs index	0.115***	-0.225***	-0.127***	-0.297***	-0.484***	-0.451***
	(0.024)	(0.016)	(0.015)	(0.016)	(0.033)	(0.022)
Parental education						
High school	0.059	-0.317***	-0.094**	-0.299***	-0.611***	-0.584***
diploma	(0.061)	(0.030)	(0.030)	(0.029)	(0.057)	(0.038)
Higher education	0.225**	-1.051***	-0.483***	-1.055***	-1.557***	-1.307***
	(0.071)	(0.043)	(0.040)	(0.042)	(0.109)	(0.064)
Migration status						
second generation	-0.864***	-0.064	-0.519***	-0.393***	-0.057	-0.551***
	(0.137)	(0.051)	(0.055)	(0.052)	(0.084)	(0.065)
first generation	-0.379*	0.120	-0.272***	0.013	0.278**	-0.170*

	(0.150)	(0.067)	(0.071)	(0.067)	(0.102)	(0.079)
Prop. of students with parents with university degree in middle school	1.620***	-3.086***	-1.467***	-4.108***	-4.893***	-5.045***
	(0.120)	(0.093)	(0.085)	(0.096)	(0.238)	(0.148)
Prop. of migrants in middle school	-0.027	0.977***	0.300*	0.515***	0.748*	1.241***
	(0.250)	(0.151)	(0.149)	(0.151)	(0.298)	(0.202)
School average test scores in Italian	-0.182	0.169*	0.091	0.316***	0.433**	0.289**
	(0.114)	(0.076)	(0.073)	(0.076)	(0.166)	(0.105)
School average test scores in Mathematics	0.414***	0.507***	0.323***	0.215***	0.348**	0.840***
	(0.089)	(0.058)	(0.056)	(0.057)	(0.125)	(0.079)
SD for School test scores in Italian	-0.102	-0.428***	-0.135	0.078	-0.193	0.287
	(0.179)	(0.113)	(0.111)	(0.113)	(0.236)	(0.151)
SD for School test scores in Mathematics	-0.105	0.323**	0.073	0.175	0.008	0.017
	(0.157)	(0.101)	(0.098)	(0.101)	(0.214)	(0.137)
School average test scores in Italian	0.247	-0.258	-0.314*	-0.170	-1.221***	-0.642**
	(0.254)	(0.162)	(0.155)	(0.161)	(0.331)	(0.217)
% of girl in the class	-3.103***	0.919***	-0.113	1.758***	-0.315	0.049
	(0.222)	(0.139)	(0.134)	(0.137)	(0.290)	(0.189)
Constant	104,604					
Observations	-142416.1					
Log-likelihood	0.218					
Pseudo R2						

Note: Traditional STEM is the omitted choice. Parental education: no High school diploma is omitted
Standard errors in parentheses. * 0.05 ** 0.01 *** 0.001

Table A2 – Impact of grades and test scores on educational choices (Marginal effects for girls)

						Parental education			
	Grade in Italian	Grades in Mathematics	Test scores in Italian	Test scores in Mathematics	Escs index	High school diploma	Higher education	Second generation	First generation
Traditional	-0.009***	0.098***	-0.010***	0.070***	0.019***	0.023***	0.085***	0.055***	0.018*
STEM	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.005)	(0.006)	(0.008)
Traditional	0.035***	-0.010***	0.029***	-0.008***	0.010***	0.007**	0.029***	-0.023***	-0.012*
Humanities	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.004)	(0.005)
Non Trad./	-0.017***	-0.005	-0.025***	-0.010***	-0.007***	-0.021***	-0.076***	0.057***	0.046***
Tech.	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.006)	(0.007)	(0.008)
Other									
Non Trad./	0.057***	-0.056***	0.065***	-0.045***	0.009***	0.031***	0.037***	-0.065***	-0.054***
Humanities	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)	(0.007)	(0.008)	(0.010)
Tech	-0.022***	0.004*	-0.013***	0.014***	-0.007***	-0.005**	-0.024***	-0.008**	0.006
STEM	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Vocat.	-0.006***	-0.004***	-0.006***	-0.000	-0.004***	-0.005***	-0.010***	0.005**	0.006***
STEM	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Vocat.	-0.037***	-0.027***	-0.039***	-0.020***	-0.020***	-0.029***	-0.042***	-0.022***	-0.011**
Other	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)	(0.003)	(0.004)

	Prop of students with parents with university degree	Prop of migrants in middle school	Average Italian test scores	Average Maths test scores	SD Italian test scores	SD Maths test scores	% girls
Traditional	0.236***	-0.062***	-0.013	-0.051***	0.021	-0.015	0.029
STEM	(0.009)	(0.016)	(0.008)	(0.006)	(0.012)	(0.010)	(0.016)
Traditional	0.127***	-0.014	-0.011*	0.008*	0.000	-0.008	0.019
Humanities	(0.005)	(0.011)	(0.005)	(0.004)	(0.008)	(0.007)	(0.011)
Non Trad./	-0.156***	0.089***	0.005	0.027***	-0.070***	0.045***	0.007
Tech. Other	(0.013)	(0.018)	(0.010)	(0.007)	(0.014)	(0.013)	(0.021)
Non Trad./	0.142***	-0.056*	-0.007	-0.006	-0.003	-0.011	-0.024
Humanities	(0.014)	(0.023)	(0.012)	(0.009)	(0.018)	(0.016)	(0.025)
Tech	-0.124***	-0.006	0.012**	-0.013***	0.015*	0.003	0.009
STEM	(0.006)	(0.008)	(0.004)	(0.003)	(0.006)	(0.005)	(0.009)
Vocat.	-0.030***	0.000	0.004	-0.002	-0.002	-0.002	-0.014**
STEM	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)
Vocat.	-0.196***	0.050***	0.010	0.037***	0.038***	-0.012	-0.027
Other	(0.010)	(0.013)	(0.007)	(0.005)	(0.010)	(0.009)	(0.014)

Note: Impact of grades and test scores is measured as an increase equal to 1 standard deviation. Parental education: no High school diploma is omitted
Standard errors in parentheses. * 0.05 ** 0.01 *** 0.001

Table A3 – Impact of grades and test scores on educational choices (Marginal effects for boys)

	Grade in Italian	Grades in Mathematics	Test scores in Italian	Test scores in Mathematics	Escs index	Parental education High school diploma	Higher education	Second generation	Grade in Italian
Traditional	0.032***	0.082***	0.021***	0.062***	0.032***	0.038***	0.135***	0.060***	0.008
STEM	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)	(0.006)	(0.007)	(0.009)
Traditional	0.028***	-0.011***	0.024***	-0.010***	0.007***	0.005***	0.022***	-0.019***	-0.011**
Humanities	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
Non Trad./	0.009***	-0.011***	0.006*	-0.030***	-0.002	-0.011***	-0.042***	0.029***	0.018**
Tech. Other	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)	(0.003)	(0.004)	(0.005)	(0.006)
Non Trad./	0.029***	-0.031***	0.029***	-0.030***	0.004***	0.010***	0.015***	-0.021***	-0.019***
Humanities	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
Tech	-0.064***	0.001	-0.053***	0.033***	-0.026***	-0.017***	-0.092***	-0.042***	0.005
STEM	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)	(0.007)	(0.007)	(0.009)
Vocat.	-0.017***	-0.011***	-0.013***	-0.006***	-0.006***	-0.008***	-0.015***	0.008***	0.008**
STEM	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Vocat.	-0.017***	-0.018***	-0.015***	-0.019***	-0.011***	-0.017***	-0.023***	-0.014***	-0.010***
Other	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)

	Prop of students with parents with university degree	Prop of migrants in middle school	Average Italian test scores	Average Maths test scores	SD Italian test scores	SD Maths test scores	% girls
Traditional	0.425***	-0.081***	-0.029**	-0.052***	0.011	-0.022	0.029
STEM	(0.011)	(0.019)	(0.009)	(0.007)	(0.014)	(0.012)	(0.020)
Traditional	0.099***	-0.009	-0.009*	0.009**	-0.002	-0.006	0.012
Humanities	(0.004)	(0.008)	(0.004)	(0.003)	(0.006)	(0.005)	(0.008)
Non Trad./	-0.046***	0.065***	-0.003	0.031***	-0.053***	0.027**	-0.005
Tech. Other	(0.009)	(0.013)	(0.007)	(0.005)	(0.010)	(0.009)	(0.015)
Non Trad./	0.071***	-0.012	-0.006	0.006	-0.007	-0.003	-0.011
Humanities	(0.006)	(0.009)	(0.005)	(0.003)	(0.007)	(0.006)	(0.010)
Tech	-0.418***	-0.002	0.039**	-0.023*	0.038*	0.015	0.020
STEM	(0.016)	(0.023)	(0.012)	(0.009)	(0.018)	(0.016)	(0.025)
Vocat.	-0.036***	0.002	0.005	-0.000	-0.005	-0.004	-0.025**
STEM	(0.006)	(0.007)	(0.004)	(0.003)	(0.006)	(0.005)	(0.008)
Vocat.	-0.095***	0.036***	0.003	0.030***	0.019**	-0.007	-0.020*
Other	(0.007)	(0.009)	(0.005)	(0.004)	(0.007)	(0.006)	(0.010)

Table A4- Predicted probabilities of school choices for various levels of test scores in Maths and Italian

		Maths test scores					
		-1.5 s.d. (wrt average)		0 s.d. (wrt average)		1.5 s.d. (wrt average)	
	Italian Test scores	Girls	Boys	Girls	Boys	Girls	Boys
Traditional STEM		**1%**	**3%**	**10%**	**11%**	**40%**	**27%**
Traditional Humanities		0%	0%	0%	0%	0%	0%
Nontrad/tech other	-1.5 s.d. (wrt average)	24%	11%	30%	12%	23%	8%
Nontrad Humanities		17%	6%	15%	2%	7%	1%
Technical STEM		**7%**	**35%**	**15%**	**57%**	**17%**	**59%**
Vocational STEM		7%	19%	4%	8%	2%	2%
Vocational other		44%	26%	26%	11%	10%	3%
Traditional STEM		**2%**	**7%**	**14%**	**24%**	**50%**	**49%**
Traditional Humanities		3%	3%	3%	2%	1%	1%
Nontrad/tech other	0 s.d (wrt average)	26%	23%	29%	17%	19%	9%
Nontrad Humanities		46%	22%	37%	9%	17%	2%
Technical STEM		**5%**	**29%**	**9%**	**41%**	**10%**	**37%**
Vocational STEM		2%	4%	1%	2%	1%	0%
Vocational other		17%	12%	8%	5%	2%	1%
Traditional STEM		**2%**	**8%**	**11%**	**30%**	**45%**	**63%**
Traditional Humanities		17%	22%	18%	17%	11%	7%
Nontrad/tech other	1.5 s.d. (wrt average)	14%	20%	15%	15%	11%	6%
Nontrad Humanities		62%	36%	52%	20%	28%	6%
Technical STEM		**2%**	**10%**	**3%**	**17%**	**4%**	**16%**
Vocational STEM		0%	0%	0%	0%	0%	0%
Vocational other		3%	2%	1%	1%	0%	0%

Table A5 – First step of IV estimation

	Test scores in Italian in Grade 8	Test scores in Mathematics in Grade 8
Female	0.133 (0.0042)***	-.169 (0.005)***
Grade in Italian	0.208 (0.0039)***	0.079 ().005)***
Female* Grade in Italian	-.0107 (0.005)**	.003 (0.006)***
Grade in mathematics	.109 (0.003)***	.334 (0.004)***
Female* Grade in Mathematics	.019 (0.005)***	-.0147 (0.006)***
Test scores in Italian in Grade 5	.327 (0.004)***	.102 (0.005)***
Female* Test score In Italian in Grade 5	-.015 (0005)***	-.016 (0.006)***
Test scores in Mathematics in Grade 5	.125 (0.004)***	.464 (0.005)***
Female* Test score In Mathematics in Grade 5	.014 (0.005)***	-.049 (0.006)***
Escs index	.011 (0.003)***	.010 (0.003)***
Parental education		
High school diploma	.029 (0.005)***	.042 (0.005)***
Higher education	.089 (0.007)***	.131 (0.008)***
Migration status		

second generation	-.037 (0.009)***	.036 (0.010)***
first generation	-0.02 (0.010)**	.016 ().012)
Prop. of students with parents with university degree in middle school	-.138 (0.015)**	-.091 (0.018)***
Prop. of migrants in middle school	.384 (0.026)***	.306 (0.030)***
School average test scores in Italian	.742 (0.012)***	-.263 (0.014)***
School average test scores in Mathematics	-.10 (0.009)***	.837 (0.011)***
SD for School test scores in Italian	.183 (0.019)***	-.045 (0.022)**
SD for School test scores in Mathematics	-.023 (0.017)	.116 (0.020)***
% of girl in the class	-.205 (0.026)***	.227 (0.031)***
Constant	-.205 (0.022)***	-.28 (0.026)***
Observations	100,022	100,022
Pseudo R2	0.5881	0.6206

Table A6 – Impact of grades and test scores on educational choices (Coefficients with IV estimation)

	Traditional Humanities	Non Trad./ Tech. Other	Non Trad Humanities	Technical STEM	Vocational STEM	Vocational Other
Test scores in	0.542***	-0.121***	0.172***	-0.335***	-0.666***	-0.449***
Italian (residuals)	(0.048)	(0.031)	(0.037)	(0.025)	(0.056)	(0.044)
Test scores in	-0.355***	-0.516***	-0.567***	-0.211***	-0.544***	-0.575***
Maths (residuals)	(0.043)	(0.027)	(0.031)	(0.021)	(0.047)	(0.036)
Female	1.083***	1.182***	2.115***	-1.000***	0.421***	1.134***
	(0.051)	(0.026)	(0.028)	(0.028)	(0.076)	(0.041)
Grade in Italian	0.831***	-0.383***	0.065*	-0.676***	-1.400***	-1.018***
	(0.038)	(0.024)	(0.028)	(0.019)	(0.058)	(0.038)
Female*Grade in	0.084	0.042	0.039	0.121***	0.524***	0.109*
Italian	(0.048)	(0.032)	(0.033)	(0.036)	(0.084)	(0.049)
Grade in	-0.686***	-0.810***	-1.050***	-0.622***	-1.408***	-1.272***
Mathematics	(0.037)	(0.023)	(0.027)	(0.018)	(0.055)	(0.038)
Female* Grade in	-0.113*	-0.187***	-0.002	-0.197***	-0.016	-0.250***
Mathematics	(0.047)	(0.031)	(0.033)	(0.034)	(0.085)	(0.050)
Female* Test scores	0.000	-0.017	-0.017	0.100*	0.132	-0.012
In Italian (res.)	(0.060)	(0.042)	(0.045)	(0.047)	(0.092)	(0.058)
Female*Test scores in	-0.111*	0.029	0.018	-0.038	0.096	-0.039
Mathematics (res.)	(0.054)	(0.036)	(0.038)	(0.040)	(0.080)	(0.049)
Escs index	0.134***	-0.247***	-0.139***	-0.317***	-0.541***	-0.500***
	(0.025)	(0.016)	(0.016)	(0.016)	(0.035)	(0.022)
High school diploma	0.050	-0.363***	-0.123***	-0.333***	-0.659***	-0.646***
	(0.062)	(0.030)	(0.031)	(0.030)	(0.058)	(0.039)
Higher education	0.190**	-1.167***	-0.565***	-1.135***	-1.689***	-1.441***
	(0.073)	(0.044)	(0.041)	(0.043)	(0.113)	(0.066)
Migration status (ref. citizen)						

second generation	-0.997***	-0.007	-0.516***	-0.339***	0.062	-0.466***
	(0.144)	(0.051)	(0.056)	(0.053)	(0.086)	(0.066)
first generation	-0.459**	0.186**	-0.239***	0.062	0.439***	-0.044
	(0.156)	(0.067)	(0.072)	(0.068)	(0.103)	(0.080)
Prop of students with	1.601***	-2.816***	-1.330***	-3.891***	-4.449***	-4.650***
Parents with uni degree	(0.121)	(0.095)	(0.086)	(0.098)	(0.244)	(0.150)
Prop of migrants in	0.129	0.811***	0.247	0.315*	0.176	0.897***
Middle school	(0.255)	(0.153)	(0.152)	(0.154)	(0.308)	(0.206)
School average test scores in Italian	0.449***	0.111	0.379***	0.036	-0.087	-0.067
	(0.111)	(0.075)	(0.072)	(0.075)	(0.166)	(0.105)
School average test scores in Mathematics	-0.082	-0.025	-0.273***	0.001	-0.232	0.158*
	(0.087)	(0.056)	(0.055)	(0.056)	(0.122)	(0.077)
SD for School test scores in Italian	0.084	-0.457***	-0.108	0.026	-0.068	0.308*
	(0.180)	(0.114)	(0.112)	(0.114)	(0.241)	(0.153)
SD for School test scores in Mathematics	-0.243	0.224*	-0.068	0.079	-0.074	-0.007
School average test scores in Italian	(0.158)	(0.102)	(0.100)	(0.102)	(0.220)	(0.139)
% of girls in the class	-0.058	-0.404*	-0.618***	-0.216	-1.359***	-0.753***
	(0.257)	(0.164)	(0.158)	(0.164)	(0.339)	(0.222)
Constant	-3.038***	0.984***	0.012	1.863***	-0.359	-0.011
	(0.225)	(0.141)	(0.137)	(0.139)	(0.298)	(0.192)
Observations	100022					
Log-likelihood	-137093.0					
Pseudo R2	0.213					

* 0.05 ** 0.01 *** 0.001

Table A7 - Impact of rankings on educational choices (Coefficients)

	Traditional Humanities	Non Trad./ Tech. Other	Non Trad Humanities	Technical STEM	Vocational STEM	Vocational Other
Female	0.780***	1.075***	1.839***	-0.869***	0.248	1.313***
	(0.184)	(0.113)	(0.118)	(0.124)	(0.268)	(0.160)
Grade in Italian	0.412***	-0.236***	0.044	-0.330***	-0.803***	-0.536***
	(0.058)	(0.040)	(0.045)	(0.031)	(0.089)	(0.061)
Female*Grade in Italian	0.145*	0.112*	0.079	0.092	0.472***	0.267***
	(0.072)	(0.053)	(0.054)	(0.057)	(0.132)	(0.079)
Grade in Mathematics	-0.433***	-0.330***	-0.532***	-0.271***	-0.655***	-0.570***
	(0.066)	(0.042)	(0.051)	(0.032)	(0.098)	(0.068)
Female * Grade in Mathematics	-0.143	-0.246***	-0.071	-0.156*	-0.042	-0.213*
	(0.082)	(0.055)	(0.060)	(0.061)	(0.149)	(0.089)
Test scores in Italian	0.468***	-0.183***	0.225***	-0.421***	-0.735***	-0.538***
	(0.066)	(0.050)	(0.057)	(0.042)	(0.082)	(0.066)
Female*Test scores in Italian	0.074	-0.030	-0.037	0.172*	0.102	0.023
	(0.078)	(0.062)	(0.065)	(0.070)	(0.130)	(0.085)
Test scores in Mathematics	-0.262***	-0.488***	-0.503***	-0.218***	-0.523***	-0.611***
	(0.056)	(0.038)	(0.043)	(0.029)	(0.072)	(0.053)
Female*Test scores in Mathematics	-0.054	0.012	0.006	-0.077	-0.015	-0.065
	(0.068)	(0.048)	(0.050)	(0.053)	(0.108)	(0.067)
Rank in Italian grades (year 7)	1.686***	0.065	0.439**	-0.790***	-0.943**	-0.702***
	(0.242)	(0.146)	(0.166)	(0.116)	(0.304)	(0.209)
Female* Rank in Italian grades (year 7)	-0.079	-0.308	-0.241	0.003	-0.255	-0.634*
	(0.319)	(0.201)	(0.207)	(0.219)	(0.466)	(0.282)
Rank in Mathematics grades (year 7)	-0.730**	-0.784***	-0.989***	-0.650***	-1.039***	-0.931***
	(0.262)	(0.153)	(0.184)	(0.122)	(0.313)	(0.225)
Female*Rank in Mathematics (year 7)	0.028	0.133	-0.043	-0.260	-0.394	-0.157
	(0.333)	(0.210)	(0.225)	(0.232)	(0.500)	(0.305)
Rank in Italian test scores at year 8	0.563*	0.071	-0.104	0.206	-0.100	0.061

	(0.226)	(0.144)	(0.166)	(0.118)	(0.267)	(0.198)
Female* Rank Italian test scores at year 8	-0.117	0.057	0.163	-0.160	0.588	-0.113
	(0.280)	(0.188)	(0.197)	(0.207)	(0.413)	(0.257)
Rank maths test scores 8	-1.079***	-0.167	-0.591***	0.057	-0.273	-0.253
	(0.227)	(0.136)	(0.158)	(0.109)	(0.256)	(0.189)
Female*Rank Mathematics test scores at year 8	0.108	-0.004	0.249	0.058	0.354	0.135
	(0.277)	(0.177)	(0.187)	(0.195)	(0.407)	(0.247)
Escs index	0.116***	-0.226***	-0.128***	-0.299***	-0.486***	-0.454***
	(0.024)	(0.016)	(0.015)	(0.016)	(0.033)	(0.022)
Highest parental education (ref. no diploma)						
High school diploma	0.046	-0.313***	-0.091**	-0.287***	-0.600***	-0.573***
	(0.061)	(0.030)	(0.030)	(0.029)	(0.057)	(0.038)
Higher education	0.204**	-1.048***	-0.484***	-1.040***	-1.548***	-1.298***
	(0.071)	(0.043)	(0.040)	(0.042)	(0.109)	(0.064)
Migration status (ref. citizen)						
second generation	-0.844***	-0.067	-0.518***	-0.409***	-0.075	-0.568***
	(0.137)	(0.051)	(0.055)	(0.053)	(0.085)	(0.065)
first generation	-0.356*	0.119	-0.272***	-0.000	0.263*	-0.183*
	(0.150)	(0.067)	(0.072)	(0.067)	(0.102)	(0.079)
Prop of students with parents with university degree in middle school	1.695***	-3.124***	-1.489***	-4.181***	-4.969***	-5.117***
	(0.121)	(0.094)	(0.085)	(0.097)	(0.238)	(0.148)
Prop of migrants in middle school	-0.142	1.027***	0.324*	0.643***	0.895**	1.360***
	(0.251)	(0.151)	(0.150)	(0.152)	(0.299)	(0.203)
School average test scores in Italian	0.001	0.157	0.074	0.265**	0.330	0.157
	(0.120)	(0.080)	(0.078)	(0.081)	(0.175)	(0.112)
School average test scores in Mathematics	0.208*	0.444***	0.200***	0.207***	0.258	0.745***
	(0.093)	(0.062)	(0.060)	(0.061)	(0.136)	(0.085)
SD for School test scores in Italian	0.089	-0.491***	-0.187	-0.028	-0.329	0.156
	(0.180)	(0.113)	(0.111)	(0.113)	(0.237)	(0.152)
SD for School test scores in Mathematics	-0.212	0.284**	0.040	0.140	0.003	0.004
	(0.157)	(0.101)	(0.098)	(0.101)	(0.216)	(0.138)

% of girls in the school	0.353	-0.290	-0.304	-0.270	-1.340***	-0.756***
	(0.255)	(0.162)	(0.156)	(0.161)	(0.330)	(0.218)
Constant	-3.444***	1.427***	0.541**	2.499***	0.960**	1.099***
	(0.268)	(0.167)	(0.169)	(0.159)	(0.337)	(0.226)
Observations	104,597					
Log-likelihood	-142057.0					
Pseudo R2	0.220					

*

Table A8 – Impact of rankings on educational choices (Marginal effects for girls)

	Grade in Italian	Grade in Mathematics	Test scores in Italian	Test scores in Mathematics	Rank Italian grades (year 7)	Rank maths grades (year 7)	Rank Italian test scores (year 8)	Rank maths test scores (year 8)
Trad STEM	-0.009**	0.075***	-0.008	0.058***	-0.015	0.113***	-0.016	0.043**
	(0.003)	(0.004)	(0.004)	(0.003)	(0.015)	(0.015)	(0.013)	(0.013)
Trad hum	0.023***	-0.010***	0.022***	-0.001	0.070***	-0.007	0.017*	-0.035***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.008)	(0.008)	(0.007)	(0.007)
Non trad/	-0.018***	-0.013**	-0.028***	-0.012**	-0.002	0.029	0.013	0.003
tech/oth	(0.004)	(0.005)	(0.005)	(0.004)	(0.016)	(0.017)	(0.016)	(0.015)
Non trad hum	0.038***	-0.037***	0.066***	-0.034***	0.088***	-0.094***	-0.007	-0.040*
	(0.004)	(0.005)	(0.006)	(0.005)	(0.017)	(0.018)	(0.017)	(0.017)
Tech STEM	-0.014***	0.007*	-0.011**	0.010***	-0.040***	-0.010	-0.002	0.022*
	(0.003)	(0.003)	(0.004)	(0.003)	(0.011)	(0.011)	(0.010)	(0.010)
Vocational	-0.004*	-0.002	-0.007***	-0.001	-0.012*	-0.009	0.007	0.004
STEM	(0.002)	(0.002)	(0.002)	(0.001)	(0.005)	(0.006)	(0.005)	(0.005)
Vocat Other	-0.017***	-0.020***	-0.033***	-0.019***	-0.089***	-0.022	-0.012	0.004
	(0.003)	(0.004)	(0.004)	(0.003)	(0.012)	(0.013)	(0.011)	(0.011)

* 0.05 ** 0.01 *** 0.001

Note: Marginal effects of other variables are similar to the ones from Table A2 and are available on request.

Table A9 – Impact of rankings on educational choices (Marginal effects for boys)

	Grade in Italian	Grade in Mathematics	Test scores in Italian	Test scores in Mathematics	Rank Italian grades (year 7)	Rank maths grades (year 7)	Rank Italian test scores (year 8)	Rank maths test scores (year 8)
Trad STEM	0.026***	0.055***	0.025***	0.052***	0.014	0.115***	-0.026	0.039**
	(0.004)	(0.004)	(0.005)	(0.004)	(0.016)	(0.017)	(0.016)	(0.015)
Trad hum	0.017***	-0.010***	0.019***	-0.004	0.062***	-0.014	0.019*	-0.035***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.008)	(0.009)	(0.008)	(0.008)
Nontrad tech/	-0.002	-0.006	0.006	-0.028***	0.050***	-0.026	-0.004	-0.007
Other	(0.004)	(0.004)	(0.005)	(0.004)	(0.014)	(0.015)	(0.014)	(0.013)
Non trad hum	0.018***	-0.023***	0.033***	-0.021***	0.053***	-0.037**	-0.017	-0.038***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.011)	(0.013)	(0.012)	(0.011)
Tech STEM	-0.032***	0.005	-0.059***	0.022***	-0.151***	-0.016	0.036*	0.053**
	(0.005)	(0.005)	(0.006)	(0.005)	(0.018)	(0.019)	(0.018)	(0.017)
Vocational	-0.013***	-0.008***	-0.011***	-0.005**	-0.013	-0.009	-0.006	-0.005
STEM	(0.002)	(0.002)	(0.002)	(0.002)	(0.007)	(0.008)	(0.006)	(0.006)
Vocational	-0.014***	-0.014***	-0.013***	-0.016***	-0.015	-0.014	-0.002	-0.008
Other	(0.003)	(0.003)	(0.003)	(0.002)	(0.010)	(0.010)	(0.009)	(0.009)

* 0.05 ** 0.01 *** 0.001

Note: Marginal effects of other variables are similar to the ones from Table A3 and are available on request.

Table A10 – Impact of comparative advantage on educational choices (Coefficients)

	Traditional Humanities	Non Trad./ Tech. Other	Non Trad Humanities	Technical STEM	Vocational STEM	Vocational Other
Female	0.913***	0.600***	1.580***	-1.518***	-0.756***	0.372***
	(0.036)	(0.022)	(0.024)	(0.025)	(0.049)	(0.029)
Diff in grade Italian-maths	0.559***	0.350***	0.585***	0.218***	0.387***	0.407***
	(0.032)	(0.019)	(0.022)	(0.014)	(0.027)	(0.022)
Female* Diff in grade Italian-maths	0.068	0.128***	0.026	0.130***	0.132**	0.125***
	(0.040)	(0.025)	(0.027)	(0.027)	(0.046)	(0.029)
Diff in test score Italian-maths	0.531***	0.464***	0.624***	0.207***	0.421***	0.521***
	(0.037)	(0.020)	(0.024)	(0.015)	(0.034)	(0.027)
Female* Diff in test score Italian-maths	0.075	-0.094***	-0.095**	-0.000	-0.106	-0.118***
	(0.046)	(0.027)	(0.029)	(0.029)	(0.057)	(0.035)
Escs index	0.133***	-0.281***	-0.159***	-0.350***	-0.599***	-0.569***
	(0.024)	(0.015)	(0.014)	(0.014)	(0.033)	(0.020)
Highest parental education (ref. no diploma)	ref.	ref.	ref.	ref.	ref.	ref.
High school diploma	0.124*	-0.506***	-0.219***	-0.498***	-0.914***	-0.882***
	(0.060)	(0.028)	(0.029)	(0.027)	(0.054)	(0.035)
Higher education	0.333***	-1.368***	-0.705***	-1.364***	-2.039***	-1.788***
	(0.071)	(0.040)	(0.039)	(0.039)	(0.106)	(0.060)
Migration status (ref. citizen)	ref.	ref.	ref.	ref.	ref.	ref.
second generation	-0.963***	0.425***	-0.182***	0.127*	0.771***	0.233***
	(0.137)	(0.048)	(0.054)	(0.049)	(0.081)	(0.061)
first generation	-0.470**	0.609***	0.062	0.528***	1.126***	0.645***
	(0.145)	(0.062)	(0.068)	(0.062)	(0.096)	(0.073)

Prop of students with parents with university degree in middle school	1.295***	-1.626***	-0.523***	-2.605***	-2.366***	-2.545***
	(0.116)	(0.087)	(0.081)	(0.088)	(0.223)	(0.133)
Prop of migrants in middle school	0.001	0.369**	-0.072	-0.126	-0.486	0.108
	(0.247)	(0.141)	(0.143)	(0.140)	(0.280)	(0.183)
School average test scores in Italian	-0.080	-0.460***	-0.317***	-0.298***	-0.699***	-0.764***
	(0.114)	(0.071)	(0.071)	(0.070)	(0.153)	(0.095)
School average test scores in Mathematics	0.512***	0.370***	0.283***	0.152**	0.064	0.497***
	(0.088)	(0.054)	(0.054)	(0.053)	(0.116)	(0.071)
SD for School test scores in Italian	0.440*	-0.904***	-0.463***	-0.360***	-0.548*	-0.239
	(0.174)	(0.106)	(0.107)	(0.104)	(0.223)	(0.138)
SD for School test scores in Mathematics	-0.403**	0.133	-0.165	-0.091	0.020	0.107
	(0.154)	(0.095)	(0.094)	(0.093)	(0.205)	(0.126)
% of girls in the school	0.126	0.047	-0.085	0.195	-0.602	-0.093
	(0.254)	(0.153)	(0.151)	(0.150)	(0.314)	(0.198)
Constant	-2.952***	1.065***	0.007	1.958***	0.179	0.236
	(0.219)	(0.131)	(0.130)	(0.127)	(0.274)	(0.171)
Observations	104604					
Log-likelihood	-155463.6					
Pseudo R2	0.147					

* 0.05 ** 0.01 *** 0.001

Table A11 – Impact of comparative advantage on educational choices (Marginal effects for girls)

	Diff in grade Italian-maths	Diff in test score Italian-maths
Trad STEM	-0.084***	-0.069***
	(0.002)	(0.002)
Trad hum	0.013***	0.015***
	(0.001)	(0.001)
Nontrad tech/ Other	0.010***	0.005*
	(0.002)	(0.002)
Non trad hum	0.057***	0.055***
	(0.002)	(0.002)
Tech STEM	-0.005***	-0.009***
	(0.001)	(0.001)
Vocational STEM	0.001*	-0.001
	(0.000)	(0.001)
Vocational	0.008***	0.004**
	(0.001)	(0.001)

* 0.05 ** 0.01 *** 0.001

Note: Marginal effects of other variables are similar to the ones from Table A3 and are available on request.

Table A12 – Impact of comparative advantage on educational choices (Marginal effects for boys)

	Diff in grade Italian-maths	Diff in test score Italian-maths
Trad STEM	-0.056***	-0.061***
	(0.002)	(0.002)
Trad hum	0.011***	0.009***
	(0.001)	(0.001)
Nontrad tech/ Other	0.014***	0.026***
	(0.002)	(0.002)
Non trad hum	0.028***	0.029***
	(0.002)	(0.002)
Tech STEM	-0.010***	-0.023***
	(0.002)	(0.003)
Vocational STEM	0.004***	0.004***
	(0.001)	(0.001)
Vocational Other	0.010***	0.016***
	(0.001)	(0.002)

* 0.05 ** 0.01 *** 0.001

Note: Marginal effects of other variables are similar to the ones from Table A3 and are available on request.

Table A13 – Impact of performance of peers by gender on educational choices (Coefficients)

	Traditional Humanities	Non Trad./ Tech. Other	Non Trad Humanities	Technical STEM	Vocational STEM	Vocational Other
Female	0.930***	1.001***	1.885***	-1.112***	0.283**	0.905***
	(0.077)	(0.043)	(0.046)	(0.047)	(0.105)	(0.062)
Grades in Italian	0.724***	-0.234***	0.126***	-0.508***	-1.039***	-0.711***
	(0.039)	(0.025)	(0.028)	(0.020)	(0.058)	(0.038)
Female*Grades in Italian	0.073	0.047	0.024	0.094*	0.429***	0.118*
	(0.050)	(0.033)	(0.034)	(0.037)	(0.085)	(0.050)
Grades in mathematics	-0.587***	-0.559***	-0.804***	-0.479***	-1.003***	-0.899***
	(0.041)	(0.025)	(0.029)	(0.019)	(0.057)	(0.039)
Female* grades in mathematics	-0.121*	-0.215***	-0.065	-0.218***	-0.115	-0.263***
	(0.051)	(0.034)	(0.036)	(0.037)	(0.088)	(0.052)
Test scores in Italian	0.629***	-0.155***	0.200***	-0.367***	-0.749***	-0.531***
	(0.042)	(0.028)	(0.033)	(0.022)	(0.048)	(0.038)
Female*Test scores in Italian	0.036	-0.030	0.003	0.116**	0.222**	-0.028
	(0.052)	(0.037)	(0.039)	(0.041)	(0.081)	(0.051)
Test scores in Mathematics	-0.442***	-0.472***	-0.604***	-0.177***	-0.544***	-0.582***
	(0.038)	(0.023)	(0.027)	(0.017)	(0.040)	(0.031)
Female*Test scores in Mathematics	-0.051	0.023	0.061	-0.057	0.052	-0.031
	(0.047)	(0.031)	(0.033)	(0.034)	(0.069)	(0.043)
Grade 7 class share of high achieving girls in maths	0.233	0.289**	0.401***	0.377***	0.632***	0.645***

	(0.158)	(0.092)	(0.110)	(0.074)	(0.158)	(0.123)
Female*Grade 7 class share of high achieving girls in maths	-0.253	0.056	-0.045	0.178	0.012	0.093
	(0.199)	(0.125)	(0.134)	(0.139)	(0.260)	(0.168)
Grade 7 class share of high achieving boys in maths	-0.057	0.447***	0.343**	0.488***	0.603**	0.690***
	(0.180)	(0.103)	(0.124)	(0.084)	(0.190)	(0.139)
Female*Grade 7 class share of high achieving boys in maths	0.004	0.044	0.072	0.105	0.165	0.146
	(0.223)	(0.141)	(0.151)	(0.155)	(0.307)	(0.189)
Escs index	0.117***	-0.220***	-0.125***	-0.296***	-0.479***	-0.445***
	(0.024)	(0.016)	(0.015)	(0.016)	(0.033)	(0.022)
Highest parental education (ref. no diploma)						
High school diploma	0.055	-0.320***	-0.096**	-0.297***	-0.611***	-0.587***
	(0.061)	(0.030)	(0.030)	(0.029)	(0.057)	(0.038)
Higher education	0.213**	-1.070***	-0.496***	-1.065***	-1.572***	-1.331***
	(0.072)	(0.043)	(0.040)	(0.042)	(0.109)	(0.064)
Migration status (ref. citizen)						
second generation	-0.862***	-0.055	-0.519***	-0.385***	-0.049	-0.543***
	(0.137)	(0.051)	(0.055)	(0.052)	(0.085)	(0.065)
first generation	-0.373*	0.122	-0.272***	0.019	0.278**	-0.160*
	(0.150)	(0.067)	(0.071)	(0.067)	(0.102)	(0.079)
Prop of students with parents with university degree in middle school	1.835***	-2.600***	-1.190***	-3.736***	-4.407***	-4.233***
	(0.100)	(0.083)	(0.074)	(0.086)	(0.218)	(0.134)
Prop of migrants in middle school	-0.281	0.323*	-0.040	0.149	0.100	0.379*
	(0.230)	(0.137)	(0.137)	(0.138)	(0.273)	(0.181)
% of girls in the school	0.250	-0.315*	-0.283	-0.110	-1.181***	-0.647**
	(0.246)	(0.160)	(0.155)	(0.159)	(0.334)	(0.215)

Constant	-3.355***	0.793***	-0.303***	1.824***	-0.708***	0.102
	(0.152)	(0.090)	(0.091)	(0.088)	(0.186)	(0.122)
Observations	104435					
Log-likelihood	-142267.1					
Pseudo R2	0.218					

* 0.05 ** 0.01 *** 0.001

Table A14 – Impact of performance of peers by gender on educational choices (Marginal effects for girls)

	Grade 7 class share of high achieving girls in maths Coef./S.E.	Grade 7 class share of high achieving boys in maths Coef./S.E.
Trad STEM	-0.043***	-0.051***
	(0.009)	(0.010)
Trad hum	-0.011*	-0.014**
	(0.005)	(0.005)
Nontrad tech/	-0.005	0.011
Other	(0.010)	(0.011)
Non trad hum	0.013	0.011
	(0.011)	(0.012)
Tech STEM	0.014*	0.011
	(0.007)	(0.007)
Vocational STEM	0.003	0.004
	(0.003)	(0.004)
Vocational Other	0.029***	0.030***
	(0.007)	(0.008)

* 0.05 ** 0.01 *** 0.001

Note: Marginal effects of other variables are similar to the ones from Table A3 and are available on request.

Table A15 – Impact of performance of peers by gender on educational choices (Marginal effects for boys)

	Grade 7 class share of high achieving girls in maths Coef./S.E.	Grade 7 class share of high achieving boys in maths Coef./S.E.
Trad STEM	-0.055***	-0.060***
	(0.010)	(0.012)
Trad hum	0.003	-0.008
	(0.005)	(0.006)
Nontrad tech/	-0.002	0.011
Other	(0.009)	(0.010)
Non trad hum	0.011	0.002
	(0.007)	(0.008)
Tech STEM	0.021	0.037**
	(0.011)	(0.013)
Vocational STEM	0.007	0.004
	(0.004)	(0.005)
Vocational Other	0.016**	0.014*
	(0.006)	(0.006)

* 0.05 ** 0.01 *** 0.001

Note: Marginal effects of other variables are similar to the ones from Table A3 and are available on request.

Table A16 – Simulation of changes in girls' test scores in Mathematics

	Observed choice			Simulated choice			% reduction in gender gap
	M	F	Gap M-F	M	F	Gap M-F	
Traditional STEM	30.40%	20.33%	10.08%	30.27%	22.90%	7.37%	27%
Traditional Humanities	3.04%	6.32%	-3.28%	3.16%	6.05%	-2.89%	12%
Non-traditional/Technical Other	13.24%	22.33%	-9.09%	13.28%	21.86%	-8.58%	6%
Non-traditional Humanities	7.53%	33.54%	-26.02%	7.69%	31.91%	-24.22%	7%
Technical STEM	36.27%	6.78%	29.49%	35.99%	7.18%	28.81%	2%
Vocational STEM	3.19%	1.48%	1.72%	3.22%	1.46%	1.76%	-3%
Vocational Other	6.32%	9.21%	-2.89%	6.38%	8.63%	-2.25%	22%

Table A17 – Oaxaca-Blinder decomposition of gender gap (F-M) in school choice

	Traditional STEM	Traditional Humanities	Non-traditional/ Technical Other	Non-traditional Humanities	Technical STEM	Vocational STEM	Vocational Other
Boys	30.64	3.07	13.15	7.49	36.23	3.17	6.26
Girls	20.45	6.33	22.35	33.51	6.73	1.45	9.18
Gender gap	-10.19	3.26	9.20	26.02	-29.50	-1.72	2.92
Explained component	3.11	0.98	-0.07	0.66	-2.89	-0.68	-1.11
Unexplained component	-13.30	2.28	9.27	25.36	-26.61	-1.04	4.03