WHY DO WOMEN CHOOSE STEM EDUCATIONAL PATH?

Giovanna Di Castro, Valentina Ferri, Salvatore Marsiglia

1. Introduction

In recent decades, women's participation in higher education has definitely increased but not in every field of study. As has happened, for example, in the STEM (science, technology, engineering and mathematics) paths where female presence remains very low.

The gender segregation in STEM education contributes to explain the persistence of gender inequalities in the labor market, and the relatively poor occupational chances compared to male ones (Barone, 2020). The study is based on longitudinal data on a cohort of Italian high schools and shows how academic strength plays an important role in determining this gap.

Taking into account 67 countries, Mostafa (2019) finds that girls are stronger in reading and boys in science and mathematics. Also in Italy we show the same pattern (Di Castro e Ferri, 2021). These differences could explain why boys are more likely than girls to choose STEM fields, even if the overall performance are similar between them. The authors conclude that students choose their field of studies basing on their comparative strengths rather than considering their absolute strengths (Mostafa, 2019).

The gender gap in the choice of scientific subjects could have many consequences, although studies have been conducted by many authors that attribute partly the gender pay gap to the choice of fields of study, Ferri et al. (2022) find that the gender pay-gap remains high excluding from the estimations the fields of studies in which female graduates are underrepresented.

The present study aims to investigate the probability that girls choose STEM fields at university compared to boys. The research focuses on the existence of a gender gap after high school degree, and analyses the variables that mostly affect this gap. In order to investigate the gender gap we use a probit model and Oaxaca blinder decomposition.

The paper is structured as follows: the second paragraph is a literature review, the third paragraph describes data and methodology, the fourth paragraph regards the results and the fifth paragraph concludes.

2. Literature

Over the past decade, women's participation in higher education has overtaken men’s. Despite this achievement, women are systematically over-represented in the humanities and social sciences fields of study, which offer relatively poor employment prospects. At the same time, women continue to be underrepresented in more competitive and rewarding sectors than the average, such as science, mathematics, engineering, and ICT (Romito et al. 2020; OECD, 2020; 2021).

According to some authors, gender segregation in higher education is recognized as a significant factor explaining the persistence of gender inequalities in the labor market, despite the reversal of the gender gap in education and higher female rates.

Barone (2020), using a set of longitudinal data on a recent cohort of Italian upper secondary school graduates, demonstrated how the motivations linked to preferences for school subjects and specific professional careers can significantly reduce gender segregation in higher education. Furthermore, a key role is played by the choice of upper secondary level school which determines the gender difference in female access to the humanities and social sciences (2/3 of the total) instead of access to engineering and ICT (only 1/3 of the total).

Some researchers who have analyzed the role of the comparative advantage of men in mathematical fields, argue that this advantage cannot be considered a sufficient explanation to understand the genesis of such models. Indeed, comparing through the Oaxaca-Blinder method the behaviors of males and females with similar mathematical results, the dominance of males in STEM fields does not change significantly (Justman & Méndez 2016). The authors also find that family socio-economic disadvantage negatively affects the choice of STEM pathways for male students, more than it does for female students.

However, other authors, who have also used the Oaxaca-blinder decomposition method, show that women expect to earn less than men and also have lower expectations about wages of average graduates across different fields.

In any case, according to the study, it does not contribute to explaining the gender gap in the choice of STEM (Osikominu & Pfeifer 2018). The students make their college decisions also due to the future earnings streams associated with the different university paths (Arcidiacono, 2012). Such evaluation certainly affects their choices but in different way between girls and boys.

Among the authors who take into account factors related to socio-economic disadvantage, Kumar e Sahoo (2021) analyze the role of social identity and find that especially females and individuals belonging to historically disadvantaged social groups are significantly less likely to study science. In the end, no single factor can explain the persistence of this model related to the differences for choosing the university path between males and females. And it is evident that various factors contribute to stimulating and strengthening this horizontal segregation. However, focusing on the elements that play a role in this dynamic can be useful in the still distant goal of reversing this trend that risks perpetuating a waste in terms of female human capital potentially linked to the STEM sectors.

3. Data, methodology and descriptive statistics

The research is based on the survey "Educational and Professional paths of upper secondary school graduates" (ISTAT, 2015). This sample survey is carried out interviewing a cohort of graduates four years after school graduation and collects information on the results of schooling path, schooling experience, the progression of studies, the entrance in the labor market and the characteristics of the occupation. We carry out two types of empirical analyses. At the beginning, we quantify the gaps in science choice between gender using a PROBIT model, then we explore these gaps using the Blinder-Oaxaca decomposition method (Blinder, 1973; Oaxaca, 1973). For what concerns the Probit model, the dependent variable is the probability to choose the STEM path, the variable of interest is "female" (the variable takes value 1 if the individual is female). The other control variables are: Failing in school (at least one year); Mother's graduation; Father's Graduation; Unemployed_Father; Unemployed_Mother; North West; North East; Center; South; Islands; Modern School; Technical School; High School Diploma with specialization in teacher training; Scientific Grammar School; Classical Grammar School; Linguistic Grammar School; Art education; High school utility for the universitary path; High school grade and Science, Technology, Engineering and Mathematics; utility for universitary path.

We also apply the OB decomposition (Blinder, 1973; Oaxaca, 1973), in order to estimate the amount of the differential between male and female STEM enrollment. Through this method we distinguish which part is the result of the differences in characteristics included in model estimations and which part remains unexplained. We estimated the threefold decomposition, dividing the differences in probability to choose STEM into endowments (Endowments, due to differences in the predictors), coefficient (C, the contribution of the unexplained component) and interaction effects (I, indicating simultaneous differences) between the two groups:

$$\text{Endowments} = \{E(XM) - E(XF)\}\ \beta M \tag{1}$$

$$\text{Coefficient} = E(XM)\ (\beta M - \beta F) \tag{2}$$

$$\text{Interaction} = \{E(XM) - E(XF)\}\ (\beta M - \beta F)) \tag{3}$$

Our reference group are females. Group differences in predictors are weighted by the coefficients of the female group to calculate the endowments effect

$$R = \{E(XM) - E(XF)\}\ \beta M + E(XM)\ (\beta M - \beta F) - \{E(XM) - E(XF)\}\ (\beta M - \beta F) \tag{4}$$

The first term of (4) is the endowment effect, and the second term is the "coefficient effect", the third is interaction effect. It is important to note that the sample of observation used to carry out the estimates are composed of those secondary school graduates that enroll into University. This occurrence could raises the issue of selection bias (Heckman, 1979) because attending University may be not random. If unobserved variables that affect the decision to carry on studying, potentially affecting women and men in a different manner (Mead, 2022), also influence the choice of field of study, results will be biased. This could be a case of selection bias because the weaker students are more likely to self-select into the group of employees and the stronger choose to continue their studies.

In order to address this issue we consider for future improvements the possibility of using the correction à la Heckman in the Oacaxa Blinder estimates.

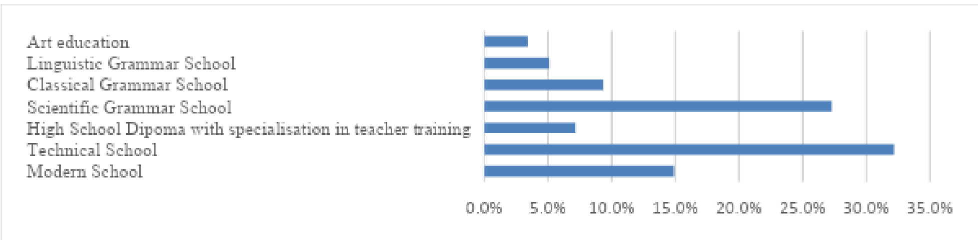
Table 1 – *Descriptive Statistics.*

Variable	mean	sd	min	max	N
Failing in school (at least one year)	0.18656	0.389565	0	1	26235
Mother's graduation	0.099869	0.299831	0	1	26235
Father's Graduation	0.114909	0.318918	0	1	26235
Unemployed_Father	0.015437	0.123284	0	1	26235
Unemployed_Mother	0.409199	0.491695	0	1	26235
North West	0.224771	0.41744	0	1	26235
North East	0.170307	0.37591	0	1	26235
Center	0.203415	0.402546	0	1	26235
South	0.277440	0.447744	0	1	26235
Islands	0.111381	0.314609	0	1	26235
Modern School	0.148856	0.355953	0	1	26235
Technical School	0.321832	0.467188	0	1	26235
High School Diploma with specialization in teacher training	0.071454	0.257587	0	1	26235
Scientific Grammar School	0.272843	0.445429	0	1	26235
Classical Grammar School	0.093369	0.290955	0	1	26235
Linguistic Grammar School	0.050842	0.21968	0	1	26235
Art education	0.033973	0.181164	0	1	26235
High school utility for the universitary path	6.575979	2.275569	0	10	11583
High school grade	76.56789	11.84886	60	101	26235
Science, Technology, Engineering and Mathematics	0.275596	0.446831	0	1	13619

Source: Authors' elaboration on ISTAT data

Table 1 shows descriptive statistics. The percentage of individuals who failed in school is 18.7%; mother and father graduation 10% and 11%. The percentage of individuals with an unemployed father is 1.5% and the percentage of unemployed mothers 41%. About the geographic distribution the table shows the same percentage of almost 40% in the North and in South+Islands. In the Center Area a percentage of around 20% is recorded.

Figure 1. – *Type of secondary school.*



Source: Authors' elaboration on ISTAT data

Regarding the different types of school, it seems to appear clearly how the Scientific Grammar School (27.3%) and Technical School (32.2%) collected the majority of the STEM students. The graduates do not consider very important the high school for the university path (65.8 of 100) and their High School grade is medium/high (75.6 of 100). They have also chosen Science, Technology, Engineering and Mathematics subjects for the 27.6%.

4. Results

In this section, we investigate the gender gap using a probit model. The table 2 shows that the gender gap in STEM choice is 14.3 p.p. The probability that women choose STEM in university path is 14.3 p.p. lower than men.

In the equation we include other control variables, we focus on some variables that are most significant as types of school, failing in school, high school grade and utility for universitary path.

Table 2 –*Probability to choose STEM in university path, probit model (marginal effects).*

	STEM_choose b/se
Female	-0.1426*** [0.0130]
Failing in school (at least one year)	-0.0425** [0.0213]
Modern School	0 [.]
Technical School	0.1008*** [0.0185]
High School Diploma with specialization in teacher training	-0.0602*** [0.0167]
Grammar School (Scientific)	0.1551*** [0.0182]
Grammar School (Classical)	-0.0104 [0.0213]
Grammar School (Linguistic)	-0.0700*** [0.0212]
Art education	-0.0378* [0.0207]
High school utility for the universitary path	-0.0083*** [0.0027]
High school grade	0.0047*** [0.0006]
N	11121

Source: Authors' elaboration on ISTAT data. Variables included: Mother's graduation, Father's Graduation, South and Islands, High school utility for the universitary path, Motivation, Reason for enrolment, Reason for choosing this University, Course choice channels, Father's job, Mother's job.

The table shows that with respect to modern school, the scientific high school increases the probability of choosing a STEM path of 15.5 p.p., such as the technical school diploma (+10.1 p.p.). For what concerns high school with specialization in teacher training and linguistic school, this kind of path decreases the probability to choose a STEM path. As the graduation grade increases, the probability of choosing a STEM path also increases. The variable High school utility for the universitary path represents the idea that the graduates have in terms of utility of high school for the universitary path. We observe that this variable is negatively correlated with the choice of STEM path.

In this section, we further investigate the gender gap using decomposition analysis. The table 3 presents the results for the gender gap: the overall gap of 18 percentage points; the decomposition highlights part of the gap which is explained by the differences in observable characteristics (endowments) 4.5 p.p.

This finding suggests that the gender gap in scientific choice is mainly driven by differences in unobservable factors. The effect of the coefficient represents 90% of the difference.

Table 3. – *Oaxaca Blinder decomposition, overall results.*

	OBeform_Tebb
	b/se
Overall	
group_1	0.3692***
	[0.0077]
group_2	0.1903***
	[0.0047]
Difference	0.1789***
	[0.0090]
endowments	0.0445***
	[0.0091]
coefficients	0.1593***
	[0.0096]
interaction	-0.0250**
	[0.0098]

Source: Authors' elaboration on ISTAT data.

As regard endowments, estimates indicate that in Italy the difference in the probability of STEM enrollment between men and women is about 18 points.

Explained variables show that High School grade affects the probability relate to the choice STEM for the women which determinate the decrease of the differential (-0.0174 p.p.) between males and females. This result seems to suggest that being a «good student» also increases the possibility for women to choose STEM paths. The data match up to numerous literature indications that show that greater self-confidence women could encourage the choice of paths instead of the others. The motivation for choosing the path «To increase future job opportunities» (0.0038 **) is significant for the endowment effect. This result, although the value is small, is significant. However, it suggests that the choice based on the more opportunities is directed towards NON-STEM pathways for women.

This consideration seems to underline how the awareness of a labor market, partly "segregated" by gender, is already rooted in the girls at the time of university choice. At the same time, the awareness of job opportunities can be also addressed by the family influence during the decisional process for choosing the university path. However, this aspect cannot be explored in this work.

If we consider high school attendance, we note that the technical and scientific school increase the difference between men and women (0.0155 *** and 0.240 ***) decreasing the women's probability for choosing STEM courses.

With regard to the coefficient effect (Table 5), the results of our analysis suggest that the gender gap in STEM choice is driven largely by other unobservable factors (about 85%).

In the unexplained part of OB decomposition, it is observed how motivation (0.0845***), the correspondence between enrollment in the university course and the real student wishes, is significant and tends to increase the gender gap increasing discrimination effects for women.

Table 4. – *Oaxaca Blinder decomposition, endowments.*

Endowments	
Failing in school (at least one year)	-0.0026*
	[0.0016]
High school grade	-0.0174***
	[0.0023]
High school utility for the universitary path	-0.0005
	[0.0009]
Motivation	0.0006
	[0.0016]
Interest in studying / in the specific university discipline	-0.0029
	[0.0022]
To increase future job opportunities	0.0038**
	[0.0019]
Family pressions o simple advices	-0.0001
	[0.0003]
To do something waiting for a job	-0.0019***
	[0.0007]
Others (Reason for enrolling in university)	0
	[0.0001]
Modern school	-0.0002
	[0.0003]
Technical school	0.0262***
	[0.0033]
High School Diploma with specialization in teacher training	0.0155***
	[0.0053]
Grammar School (Scientific)	0.0240***
	[0.0030]
Grammar School (Classical)	0.0033
	[0.0021]
Grammar School (Linguistic)	0.0066**
	[0.0034]
Art education	0.0013
	[0.0008]
North West	0.0003
	[0.0003]
North East	-0.0001
	[0.0002]
Centre	-0.0001
	[0.0002]
South	-0.0002
	[0.0002]
Islands	0.0002
	[0.0003]
N	11121

Source: Authors' elaboration on ISTAT data.

The analysis shows that the gap is widened for individuals living in the northern areas and in particular in the North-West (0.0124***) and in the North East (0.0063*). On the other hand, in the Center, in the South and in the Islands there is a negative coefficient and in particular in the South it is significant (-0.0118***). This latest evidence shows how living in the South supports a smaller gap between men and women in the STEM choice.

We deepened these issues with descriptive analysis and about the University enrollments by geographical area with particular reference to the total of the two genders (males and females), we verify that in the northern and central regions the percentage of women enrolled is greater than what we have registered in the South and in the Islands, with a significant difference of almost 8/9 percentage points. In the same way, men enrolled in the South are also lower than the percentage of those enrolled in the North.

Table 5. – *Oaxaca Blinder decomposition, coefficient effect.*

Coefficients	
Failing in school (at least one year)	0.0004
	[0.0039]
Mother's Graduation	-0.007
	[0.0049]
Father's Graduation	-0.0131**
	[0.0055]
High school grade	0.2790***
	[0.0609]
High school utility for university path	0.0944***
	[0.0255]
Motivation	0.0845***
	[0.0277]
Interest in studying / in the specific university discipline	0.0054
	[0.0193]
To increase future job opportunities	0.0063
	[0.0177]
Family pressions o simple advices	-0.0019
	[0.0012]
To do something waiting for a job	-0.0045***
	[0.0013]
Others (Reason for enrolling in university)	0.0010**
	[0.0005]
Modern school	0.0011
	[0.0014]
Technical school	0.0425***
	[0.0067]
High School Diploma with specialisation in teacher training	-0.0018*
	[0.0010]
Grammar School (Scientific)	0.0151
	[0.0101]
Grammar School (Classical)	-0.0037
	[0.0030]
Grammar School (Linguistic)	-0.0005
	[0.0011]
Art education	-0.0009
	[0.0008]
North West	0.0124***
	[0.0042]
North East	0.0063*
	[0.0033]
Centre	-0.0039
	[0.0038]
South	-0.0118***
	[0.0042]
Islands	-0.0019
	[0.0022]
_cons	-0.3823***
	[0.0780]
N	11121

Source: Authors' elaboration on ISTAT data.

Analyzing the data relating to STEM / NON STEM enrollments in the geographical macro-areas, it comes out how there are differences regarding the percentage of women who choose STEM subjects, which would seem to be uniform in the various macro-areas with a small difference in the South, where they represent 20.63%, while it drops to 17.98% in the North West and 17.98% in the North East, and then rises to 19.19% in the Center.

For this reason, we have been able to verify that the people who study in the south universities are overall less and this applies to both males and females, but the women who choose STEM in the South are also slightly higher.

The high school grade and high school utility for the university path play a crucial role for girls. The results confirm the importance of high school grade as a driver for the choice of STEM paths in the unexplained part.

This result seems to highlight that this factor plays a crucial role not only for female students, in the possibility of choosing scientific-mathematical paths as we have seen in the endowments part, but also in terms of discrimination.

It therefore seems that only high-performance students can invest in these paths, but this seems to be more true for girls than boys, who enroll in higher percentages regardless of the final grade achieved. This result, confirmed in the literature by other studies, as previously highlighted, constitutes an indirect confirmation for student females who need external and objective elements, as a high school grade, to make decisions on their paths in those disciplines where they are more insecure.

5. Conclusions

The gender segregation in tertiary education due to STEM fields constitutes a relevant area of study with important implications for women employment.

The under-representation of women in mathematical and scientific fields is seen to be a restriction on economic growth, especially within the European Union, as well as these fields seem to be the most requested in the near future, and where the female presence is still insufficient. Analyzing the Italian case through Istat 2015 data, we highlight how the probability of choosing STEM degree courses is significantly lower for women, despite the higher share of high school graduates and their greater presence in tertiary education. At the endowment level, a high grade increases the probability that women can choose STEM paths, underlining the role of positive strengthening that the grade can represent for the female gender, especially in fields where girls are known to perform lower than boys (Di Castro, Ferri 2021). Also Perez-Felkner et al. (2017) analyzes how beliefs about math skills in secondary school, which are known to change according to gender, influence university choice. In their studies, the authors show how girls' negative perceptions of their own mathematical competence push them away from scientific careers. However, in the decomposition we observe how a "High school grade" and the "High school utility for the university path", constitute inexplicable factors of "discrimination" among the girls and increase their probability of not choosing those paths. It would therefore seem that only high-performance male students can invest in these paths, which suggests the presence of unexplained discrimination.

In fact, a problem of "better performance" (high school grade) seems to emerge in the study as an obstacle to equal access to STEM courses. It can discourage girls from investing in such courses and penalize them compared to male colleagues due to multiple deeply rooted factors. Finally, through the Blinder-Oaxaca decomposition method (Blinder, 1973; Oaxaca, 1973) the findings suggest that the characteristics observed between males and females represent only 15% of the gender gap in choice STEM. While, the large part of the differential (about 85%) is the consequence of other unobservable factors and therefore by "discrimination" factors. Moreover, this factors have a much higher weight than the differences in the characteristics that can be observed between boys and girls.

All these findings seem to suggest that female students choose the field of path university underestimating their chances of success.

That would be the result of reasons due to insecurity about their own abilities in the scientific and technological fields and to less conscious reasons more deeply rooted that reinforce the existing segregation, penalizing them compared to their male colleagues.

References

ARCIDIACONO P., HOTZ V.J.,KANG S. 2012. Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals, *Journal of* Econometrics, Vol. 166, No.1, pp.3-16.

BARONE C., ASSIRELLI G. 2020. Gender Segregation in Higher Education: An Empirical Test of Seven Explanations, *Higher Education*, Vol 79, No. 1, pp 55-78.

BLINDER A. S. 1973. Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human resources*, Vol. 8, No. 4, pp. 436-455.

DI CASTRO G., FERRI V. 2021. What Matters Most to Math Gender Gap? Evidence from PISA Data in Italy (2021), in Proceedings of the 2nd International Conference of the Journal Scuola Democratica “Reinventing Education”, VOL. 2, Learning with New Technologies, Equality and Inclusion, pp. 1015-1026.

HECKMAN J.J. 1979. Sample Selection Bias as a Specification Error, *Econometrica*, Vol. 47, No. 1, pp. 153-161

ISTAT. 2015. Educational and Professional paths of upper secondary school graduates, from https://www.istat.it/it/archivio/35940

JUSTMAN M, MENDEZ S. J. 2016. *Gendered selection of STEM subjects for matriculation.* Working Paper No. 10/16, Melbourne Institute of Applied Economic and Social Research.

KUMAR A, SAHOO S. 2021. *Social Identity and STEM Choice: Evidence from Higher Secondary Schooling in India*. No. 900. Global Labor Organization (GLO).

MEAD D. 2022. The Gender Gap in University Enrolment: Evidence from Subjective Expectations, *Education Economics*, Vol. 31, No. 1, pp. 54-76.

MOSTAFA T. 2019. Why don’t more girls choose to pursue a science career? PISA in Focus, No. 93, OECD Publishing, Paris.

OAXACA R. 1973. Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review*, Vol. 14, No. 3, pp. 693-709.

OECD 2020. *Education at a Glance 2020: OECD Indicators*, Paris: OECD Publishing.

OECD 2021. *Education at a Glance 2021: OECD Indicators,* Paris: OECD Publishing.

OSIKOMINU A., PFEIFER G. 2018, *Perceived wages and the gender gap in STEM fields.* Discussion Paper 11321, IZA.

PEREZ-FELKNER L., NIX S., THOMAS K. 2017. Gendered Pathways: How Mathematics Ability Beliefs Shape Secondary and Postsecondary Course and Degree Field Choices, *Frontiers in Psychology*, Vol. 8, pp. 386 .

ROMITO M., GEROSA T., VISENTIN M., MARIA G. 2020. Gender Differences in Higher Education Choices. Italian Girls in the Corner? *The Education of Gender*, Vol 61.

SUMMARY

Several studies have highlighted the greater female propensity in the university choice of humanities and non-STEM (Science, Technology, Engineering and Mathematics) studies, contrary to what happens to male students. The persistence of this gender gap in educational choices affects the subsequent female under-representation in the technical-scientific fields, which are most requested and remunerated in the labor market.

Our contribution intends to investigate the probability for a student, at the end of secondary school, to undertake a technical-scientific tertiary education path. This probability is estimated through probit regressions by using the database ISTAT 2015 graduated from High School and it is calculated by taking into account first of all the gender variable and then the family and socio-economic context.

We also investigate the role of the geographical dimension in order to understand if it affects educational gender segregation. The work also intends to offer useful elements for understanding the multiple determinants that contribute increasing gender inequalities in educational models.

Giovanna DI CASTRO, Istituto Nazionale per l'Analisi delle Politiche Pubbliche (INAPP), g.dicastro@inapp.org

Valentina FERRI, Istituto Nazionale per l'Analisi delle Politiche Pubbliche (INAPP), v.ferri@inapp.org

Salvatore MARSIGLIA, Istituto Nazionale per l'Analisi delle Politiche Pubbliche (INAPP), s.marsiglia@inapp.org