UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

Dottorato di ricerca in “Lavoro, Sviluppo e Innovazione”

Ciclo XXXV

Gender inequalities in education & work: determinants, interactions and policies.

Candidata: Tasselli Chiara

Relatrice (Tutor): Prof. Tindara Addabbo

Coordinatrice del Corso di Dottorato: Prof. Tindara Addabbo

INTRODUCTION

Reducing the gender gap in education and employment is one of the priorities of the 2030 Sustainable Development Goals (SDGs) of the United Nations agenda for the coming years. The SDG 5 “Achieve gender equality and empower all women and girls” recognises the importance of gender equality, not only for ethical purposes of equity but also for economic growth.

“*Gender equality is not only a fundamental human right, but a necessary foundation for a peaceful, prosperous and sustainable world.”*

The current dissertation focus is on the determinants of gender inequalities pre and in the labour market. The essays are mainly concentrated on the pre-labour market discrimination with reference to the gender inequalities in high school education by area and their determinants and on the impact of education methodologies improving soft skills on the learning process by gender.

The first paper analyses the main determinants of university choice by using data from an experimental project in a set of Northern Italy high schools. Data have been acquired by submitting students a questionnaire in which the Implicit Association Test (IAT) is also included to measure unconscious and/or unobservable stereotypes. The model applied is based on Eccles’ Expectancy Value Model in which choices and performances are strongly shaped by two psychological-cognitive dimensions which are in turn sensitive to the level of self-stereotype.
Significant correlations between self-stereotype level and university choice have been detected. Stereotypes have also been found to be strongly associated with self-confidence, anxiety, fear, effort and the value students place on the subject and that affects choices.

The second analysis going in deep into the explanation of the first one disentangling the mechanism that through stereotype presence leads to bias choices. According to Eccles's model, a generalized structural equation model (GSEM) is specified to disentangle the relationships between stereotypes, choice drivers, and actual choices. IAT results confirm - especially for females - a nonnegligible presence of implicit stereotyping. GSEM application reveals that stereotypes in females negatively affect almost all subjective values associated with the mathematical and economic domains. The only one that seems to affect female regardless of the level of the stereotype are the costs such as anxiety, nervousness and belief of not succeeding. In turn, indicators of intrinsic motivation and ability self-concept influence the gap between economic-mathematical and humanistic skills meanwhile the sense of belonging, the extrinsic and intrinsic motivation boost enrolment intention towards universities of economics, maths and statistics.
The third essay evaluates the gender impact of the implementation of the Team-Based Learning (TBL) teaching methodology on the student’s achievement in a course on macroeconomics in a bachelor’s degree course. TBL is a teaching methodology that has been found by the literature to

improve problem-solving and team-based work. Classes show an overrepresentation of men and the quantitative content of the course exposes women to lower performance in the exams. Therefore a test on the gender impact of TBL on students' achievement in macroeconomics is relevant also for policies suggestion.

A positive and significant association between attending TBL's lessons and macroeconomic performance appears and female students benefit the most from participating in TBL: they significantly increase their chances of passing the exam and - once they pass – they do achieve a stronger improvement in their grades.

Finally, the last paper focuses on analysing gender pay differentials by using administrative data at the firm level. The sample is made of firms that voluntarily underwent a process of gender equality evaluation in Italy in 2019-2023. By applying the Blinder-Oaxaca decomposition we investigate whether this difference can be attributable to gender differences in workers' actual characteristics or whether could be imputable to the return of the observable variables. The analysis also allows us to check each observable variable's weight in explaining the wage differential and the different components - returns and characteristics.

Attention is provided to indirect discrimination starting from the choice of the university in the first two essays, or the choice of whether to work (and if so the number of hours) in the fourth article. Other sources of indirect discrimination are changes in the ability self-concept or beliefs that beside influencing choices (1st essay) can even go so far as to bias performance itself (the 3th essay shows how the TBL practice is able to change the way students approach the exam). Finally, another, more objective form of indirect discrimination is noted in the fourth essay with the close link between hourly wages and working time spent. This affects indirectly women's wages as they are more likely to work a lower number of hours.

The Ph.d candidate has been involved in the whole processes of the research projects related to different essays in the thesis. These research products, in addition to all being the result of primary data collection which starts from the design of the questionnaires and the relevant variables, entailed an active involvement of the Ph.d candidate in the data collection environment and in associated activities. The first two essays have been associated with the organization of an orientation project in high schools on the university choice, the third with the very experience of tutorship in the macroeconomic course and involvement in the TBL implementation, and the last one with a project aimed at guiding business organisations toward greater gender equality in their workforce.

Special attention is also provided to the impact on the observed inequalities of mental and psychological constructs: stereotypes, expectations, and choices that the literature is increasingly considering as conducive to gender inequalities and to (pre)labour market discrimination.

On the whole, the thesis provides evidence of gender disparities in both education and employment and suggests focusing on mitigating entrenched social beliefs as the gender-field of studies association or the patriarchal view of the society which sees women as the principal housekeeper. The main reference policies concern information for both society but also for women to increase awareness of the consequences of their choices. But also normalize "*outside the box*" choices with role models, engaging girls in quantitative and scientific pathways (and the careers that are related to them), encouraging the role of the father as a parent and the sharing of care duties in the household.

Other policies suggestions related to the results of the research can include:

- regulations that facilitate female enrolment in the quantitative academic paths (minimum percentage of females, tax reduction for female students or incentives for school/university according to the share of females, classes | courses | summer or winter school dedicated exclusively to women) to contrast the current association by gender and field of study.
- Increase paternity leave, nursery and child support services together with tax relief (or financial aid) for families who hire nannies to lower women's obligations and subsequently their reservation wage. Finally, also useful to boost women’s engagement in work, tax relief for businesses offering welfare childcare services (for all parents regardless of gender) or having policies for the retention (or reintroduction) of mothers after childbirth.

Contents

ESSAY N°4

ESSAY N° 1

Gender inequalities in college major choices. The role of explicit and implicit stereotypes.

Chiara Tasselli[a] , Tindara Addabbo[b], Chiara Strozzi[c]

a University of Modena and Reggio Emilia, Marco Biagi Foundation
Email: chiara.tasselli@unimore.it

b University of Modena and Reggio Emilia, Gender Equality Observatory, Marco Biagi Foundation
Address: Viale Berengario 51, 41121, Modena, Italy
Email: tindara.addabbo@unimore.it

c University of Modena and Reggio Emilia, Department of Economics Marco Biagi, and IZA
Address: Viale Berengario 51, 41121, Modena, Italy
Email: chiara.strozzi@unimore.it

JEL codes: J16, J71, I24, A21, C93, D91

ABSTRACT

College major choices are highly segmented by gender and generate persistent inequalities in the labour market at the disadvantage of women. The literature has underlined the role of gender stereotypes on college major choices and the costs both at individual and society level of the resulting mismatch of talents and wrong fields of education choices. Determinants of college major choices from a gender perspective are at the core of this paper which has also a special focus on the impact of gender stereotypes (implicit and explicit) and on the choice of economics as a college major.

Microdata are generated by a field experiment involving different high schools in the last year of attendance when students are close to the college major choice. The schools are located in two Northern districts of Italy a country characterized by a very high gender gap in the labour market at the disadvantage of women. Activities include role models, board game addressing gender stereotypes in professional choices and documentaries highlighting the impact of gender stereotypes in the labour market. A questionnaire including Implicit association tests has been submitted to the sample exposed to the treatment and to the control groups to measure gender stereotypes, individual and household characteristics and the impact of the treatment on college major choices and their determinants.

We find evidence of gender stereotypes among students, and – differentiating by gender – female students appear to have a higher level of implicit stereotypes than males. Female students participating in the experiment appear to increase their awareness concerning college major choice and their propensity for STEM and for finance over marketing within the economics courses whereas male students taking part in the activities appear to be more interested in pursuing higher education instead of looking for a job, are more uncertain about their future field of studies and increase their interest in humanities while showing a higher propensity for marketing instead than finance when compared to the male students' control group.

1. INTRODUCTION[1]

Differences in college major choices by gender are at the origin of the widespread underrepresentation of women in math-intensive majors and doctoral courses (OECD, 2019, 2021; European Commission, 2021). While higher earnings and better career opportunities appear to be related to male-dominated majors such as hard STEM and finance, female-dominated majors (education, humanities and a set of social sciences) are characterized by lower work opportunities and lower salaries (Carlana & Corno 2021, Carlana 2019, Cimpan 2020, Goldin 2015, Ingellis et all 2018). Thus, persistence in gender differences in fields of study is bound to be reflected in persistence in gender inequalities in the labour market (OECD, 2021).

Understanding the origins of gender differences in college major choices is therefore crucial to evaluate the existing unbalances in the labour market, leading to key effects both at the individual and at the societal level. For individuals, the first important consequence of a wrong university choice is that it generates costs (Carlana & Corno 2021), ranging from having a lower performance at work, quitting work, or even not fitting properly into future employment. For society as a whole, wrong college major choices could cause a mismatch of talents and a loss in human capital, thus potentially reducing economic growth and development.

The aim of this paper is to investigate college major choices from a gender perspective, with a special focus on how gender stereotypes may affect the preferences about the chosen field of study. A special focus of this contribution will be devoted to the choice of economics as a college major, which appears to be gender unbalanced (Paredes Fuentes et al.2021; Avilova and Goldin, 2018), especially for highly quantitative economics degree programs such as finance (Bertocchi et al. 2022); the gap is mitigated by the business and marketing degree, which is characterized by a lower mathematical intensity and a higher percentage of females (Megalokonomou et al. 2021).

To analyse the determinants of college major choices, we perform a field experiment in a sample of high schools in two Northern districts of Italy by analysing the reported preferences of students belonging to the last year of studies, which is the period immediately preceding college major choice. The choice to focus on Italy is connected to the very high gender gap in the labour market at the disadvantage of women. Indeed, in 2021 (the year when the field experiment took place) in Italy, the average male employment rate in the 15-64 age group was 67% against 73% in EU-27, while the average female employment rate was 49% against 63% in EU-27 (Eurostat, 2022). This

[1] Funding from the Gender Stereotypes and Education Gaps in the Economics Field (GSEGEF) FAR 2019 University of Modena and Reggio Emilia Interdisciplinary Research Fund project - is gratefully acknowledged. We are grateful to Professor Carlo Tomasetto for his stimulating comments on the design of the experiment and of the IAT structure and to the role models and the research assistants who participated in the project, as well as to the high school students and the institutes involved in the experiment.

gap is also visible in the Gender Equality Index, where the index dimension concerning work classifies Italy at the last position within the EU countries (EIGE, 2021). Moreover, although in Italy females graduate with higher grades have more continuity in their studies (Almalaurea 2022), once entered the labour market, they suffer from horizontal and vertical segregation (sticky floor and glass ceiling).

The gender stereotypes we focus on are both explicit and implicit. Explicit stereotypes are self-reported, conscious beliefs on gender differences in interests, attitudes and abilities in different cognitive areas; implicit stereotypes are automatic beliefs about the association between gender and a given field of education (De Gioannis, 2022a; Martin & Dinella, 2001).

The paper contributes to the literature on the determinants of college major choices in four ways: by focusing on Italy; by analysing the preferences of students in a moment when they are mostly close to college major choice; by distinguishing the role of explicit and implicit gender stereotypes on college major choices; by studying the impact of gender stereotypes on college major choice differentiating by gender, with a special focus on the economics field of study.

Our results show a relevant degree of gender stereotypes among students, and – differentiating by gender – females appear to have a higher degree of implicit stereotypes than males. Gender stereotypes have an important effect both on the determinants of the choice of college major and on the college major choice itself. Females with high gender stereotypes feel less confident in themselves and are not attracted by economics and mathematics. At the same time, they are less likely to enrol in economic degrees (preferring humanities careers) and assign less importance to economics for their future academic/professional/everyday life. For males, the effect is the opposite: the higher the degree of stereotype, the higher their level of confidence in economics and mathematics. Choice determinants such as subjective task value and expectation of success are as well differentiated by gender, being strictly correlated with university choices (positively to economics and STEM and negatively to humanities). In addition, the role of choice determinants in enrolment decision-making is higher than the role of ability and skills. As to treatment effects, treated female students appear to increase their awareness concerning college major choice and their propensity for STEM bachelor courses and for finance over marketing within the economics courses. Treated male students appear to be more interested in pursuing higher education instead of looking for a job, are more uncertain about their future field of studies and increase their interest in humanities courses. In the field of economics courses, treated males show a higher propensity for marketing instead of finance when compared to the control group.

The paper is structured as follows. Section 2 reviews the relevant literature. Section 3 describes our field experiment in high schools. Section 4 outlines the sample identification and the data. Section 5 presents the results. Section 6 concludes. The Appendix contains additional figures and tables, together with a specification of how we performed the test for the existence of implicit stereotypes.

2. LITERATURE REVIEW

The determinants of uneven gender distribution across fields of study have been largely analysed in the literature, leading to the conclusion that the factors that influence college major choice are complex and multidimensional (Patnaik et al., 2020). Psychological and cultural factors have a great role in this context and manifest themselves both at home and at school through the negative (explicit and implicit) gender stereotypes proposed by family, teachers, and peers.

Carlana and Corno (2021) show that parents affect the children's field of study either by imposing direct restrictions on their choice sets or by indirectly influencing their behaviour through recommendations, while several additional contributions highlight the role of parents on children's preferences and economic/educational decisions (Doepke et al., 2019; Lizzeri and Siniscalchi, 2008; Giustinelli, 2016). Moreover, the expectancy-value theory (EVT) of academic motivation (Eccles et al., 1983; Jacobs et al., 2005) points out that parents are a major environmental influence on the development of children's self-perception of ability in different academic domains. In particular, parents' evaluations have been found to affect the association between performance-related indicators (such as teacher's ratings) and children's self-perception of ability (e.g., Frome & Eccles, 1998; Tiedemann, 2000). As regards the role of teachers, Carlana (2019) and Lavy (2008) provide evidence that teachers' stereotypes induce girls to underperform in math, develop lower self-confidence, and self-select them into less demanding high schools. As to the role of peers, Carlana and Corno (2021) and Booth et al. (2018) show that girls do not choose math to avoid interactions in male-dominated contexts (thus perpetuating gender segregation by avoiding enrolling in a field in which they know they will be rounded by the opposite sex), Shan (2020) adds that the most likely to drop out from math-related fields are the highly skilled female and Astorne-Figari and Speer (2019) point out that students switch to majors where their gender is more represented. In addition, Feld e Zolitz (2022) find evidence that students' educational choices and labour market outcomes are affected by the proportion of female peers, showing that women who are randomly assigned to a higher proportion of female peers in their studies are more likely to choose female-dominated majors (like marketing) and less likely to choose a male-dominated major (like finance).

Together with parents, teachers and peers, also role models have been shown important for college major choices. Indeed, several studies demonstrate the positive impact of female role models on pushing female students through typical male-dominated paths, both directly by affecting their choice (Porter and Serra 2020; Breda et al. 2020) and indirectly by affecting their ability or probability to graduate (Porter and Serra 2020; Carrell et. al. 2020). Redmond et Gutke (2019) demonstrate the effectiveness of one-to-one mentorship relations in supporting females in STEM;

Jethwani, et al. (2017) tests the effectiveness of live tutoring from college and graduate students together with site visits/field trips to organizations where female role models showed how they work in generating interest in female students for cybersecurity; Stoeger and al. (2016) prove the effectiveness of the E-mentoring for women in STEM with a female mentor; Merritt et al. (2021) found an increase in science identity among adolescent girls through science workshops where female professionals in STEM talked about how they became interested in STEM.

A related branch of literature such as research in developmental and educational psychology has addressed the issue of choice drivers offering a holistic view of choice determinants, jointly considering dimensions such as motivation, beliefs, values, and goals. In particular, according to Eccles and Wingfield's model (2002), the main determinants of choices (and of performance) are:

i. *Expectations of success* (I can do It?). The expectations of success include both the observed ability and the personal prejudice about one's own abilities. That latter refers to the self-perception of being able to achieve the goal, namely the ability's self-concept. According to the literature, this second component is based on personal perception, and it is a strong determinant of course enrolment and occupational aspirations choice (Bandura 1997, Bandura et al. 2001).

ii. *Subjective task value* (Why do It?). Rokeach M. (1979) defined values as a set of stable and general beliefs about what is desirable, pointing out that beliefs are formed in the individual's basic psychological needs and sense of self and are influenced by societal norms. In other words, the *subjective task value* is a multidimensional construct which includes all dimensions that motivate (engage) individuals by influencing both the attractiveness of some goals and the thought that achieving those goals is something that should be done (Feather 1988, Feather 1992). It contains dimensions such as intrinsic value (I like it), attainment value (it makes me feel accomplished), utility (I value it important for my future goals (Eccles, 1983)), costs (anxiety, fear of failure, effort) and sense of belonging.

These dimensions are interacting blocks and develop together with the external stimuli which derive from the environment which the individual belongs to and that surround pupils from the very beginning of their lives. These external stimuli can be otherwise defined as "environmental factors" and can be identified with family, teachers, peers, cultural roles and role models. All together these factors contribute to determine pupils' future choices, with the risk of conveying gender stereotypes in the choice of the field of study. The influence of surrounding beliefs, the lack of role models (Carrell et. Al. 2010) and the lack of representation (Porter and Serra 2020) could indeed boost the *self-stereotype* and affect the association between gender and fields. Thereby, perceptions and *stereotypes* can be more relevant than interests and skills in college major choices, potentially distorting both the *expectations of success* and the *subjective tasks value*, and in turn the process that guides students' choices. In conclusion, in line with Eccles and Wingfield (2002)'s model, gender stereotypes would indirectly influence educational choices by biasing expectations of success, perceived usefulness of the course, interest in the subject, identification with the course field and the cost of choosing one course over another.

3. EXPERIMENTAL DESIGN & DESCRIPTION

The project involves students belonging to the last year of high school (fifth-grade classes)[2] in different high schools in the provinces of Modena and Reggio Emilia. The choice to focus on that sample of students is because during that year students must choose whether to enrol on universities and which college majors to choose.

As the best allocation between treated and controls is the one which identifies as a comparison group those who in the absence of the program would have had similar results to those who were exposed to the program (Duflo 2004), classes were separated into the factual and the counterfactual groups considering the institute and the various curricula so that the treated and the control groups were as similar as possible for composition and environment.

Computer-administered questionnaires have been submitted to both treated and not treated groups of students with the time sequence shown in Figure 1. At first, an ex-ante questionnaire was implemented on all students, then the treated participated in three activities (namely and in chronological order: a board game, short films projection and meeting with a role model) while the control attended classes as usual without any change in their timetable, and finally, all pupils fulfilled an ex-post questionnaire.

Figure 1 – Timeline of the field experiment

The details of the three activities are the following:

- **Board game "Free to Choose"** [time: 90 minutes]**:** students were grouped into small groups (no more than six components) and played a game aimed at changing their self-awareness regarding gender issues and gender stereotypes influencing education and career choices.

  Free to choose is a board game created with the cooperation of six southern European countries (Cyprus, Italy, Portugal, Slovenia and Spain) with funds from the European REC - Rights Equality

[2] There is only one exception where a fourth class was included as a counterfactual because it was the one in the institute with the most similar characteristics to the treated one.

and Citizenship Programme. Its conceptualisation is based on the scientific evidence arising from the explorative report by Ingellis et al. (2018) which aimed at producing evidence-based knowledge on which to establish the Free to Choose (FtC) project. Moreover, the game won the prestigious recognition of best practice in fighting gender stereotypes from the European Commission at the Information and Networking Meeting of the REC - Rights Equality and Citizenship projects dedicated to women's empowerment and combating gender-based violence (Bruxelles).

This board game is dedicated to people aged between 16 and 29 in the transition phase of their lives. In the game, points are earned by guessing professions and objects of the respective ones chosen by the participants and associated with a random superhero. The fact that players must guess "without knowing" brings out mental processes full of stereotypes, this is also accentuated by the relaxed environment of the game in which players are free to express themselves. Stereotypes also emerge from the counterpart that needs to be guessed: to give an example, to make people guess positions of power they tend to choose items such as ties rather than heels, which are instead associated with teaching and the cashier profession.

Hence the role of staff as moderators was very important: they must guide and capture the flow of the stereotypes and let them emerge in the final debriefing: for that reason, they have received specific training and studied deeply the *Trainer Handbook* which contains the guidelines for conducting the game and the debriefing.

- **Short films** from ***Short on Work*** [time: 90 minutes]: students watched and discussed three short films on gender stereotypes.

  Short on Work is an international competition for short videos on the work reality nowadays, conceived and implemented by the Marco Biagi Foundation as part of the PhD program in "Labour, Development and Innovation" at the University of Modena and Reggio Emilia. The purpose of the competition is to promote and collect audio-visual works on labour and to develop an international audio-visual archive on representations of labour today, to be used for educational and research purposes. *Short on Work,* like *Free to Choose*, is also founded on a strong theoretical and methodological framework (Capalbi, 2015). In this archive, many short films are devoted to the relationship between gender and work and the pervasive presence of stereotypes: an overview of which is provided by Capalbi and Piscitelli (2020).

  The activity in the classes used short films as a tool to reflect on the persistence of gender stereotypes and professional typecasting in representations to trigger critical reflection and class debate. Again, also in this session, the role of the staff as moderators is the key factor in the treatment quality.

- **Interactive meeting with female *role model*** [time: 90 minutes]: students were exposed to a direct contact with a female young professional engaged in an economic-financial and/or highly quantitative field who has enjoyed an economic and highly quantitative university career (role model).

The role model - told students her story (-professional life, personal achievements, academic and educational experience). The expository style was informal and designed for the young audience and presentations included experiences abroad, early college years, and current work. The selection of the role model is very important because what must emerge from her talk is that she is talented in her work, she loves it and she is comfortable in that environment, and she is happy with her path and her past choice. With the right role models, according to the literature reported in Section 2, female students can potentially overcome the stereotype problem by understanding that quantitative subjects and finance are not exclusively for "men". During the one-hour meeting, the Role Model will tell her story and interact with students by answering their questions.

The activities took place in the class environment and were mainly conducted by the same people except for the interaction with the female role model. The staff who delivered the interventions in the classes were gender-balanced and consisted of a PhD student in “Labour, Development and Innovation” and a research fellow at the Interdepartmental Center on Digital Humanities, University of Modena and Reggio Emilia (DHMORE). Random events[3] cause class meetings to sometimes be held virtually or to be subject to variations in the staff and a descriptive table of the final implementation can be found in table A4 in the Annexes.

Control students, on the other hand, continued to attend school without any change in their timetable.

To avoid interference in the outcome, none of the students is aware that it is involved in a project aimed at deconstructing gender stereotypes. They believe that they are participating in a generic orientation project.

Slots for the administration of the questionnaires were agreed upon in advance with the teachers as part of the activity and usually took place on the same day as the treated class and its corresponding counterfactual class.[4] The completion of the questionnaire in the classroom was very important to us for the accuracy and quality of the data, for the following reasons:

- fewer misunderstandings about question compilation due to the presence of staff and professors
- more effort and concentration from the students
- little interference from the external environment, same surroundings and device for all respondents in both pre and post questionnaires

[3] Random events are mainly linked to health problems and the pandemic situation.
[4] In some cases - happened that one or both questionnaires couldn’t take place exactly the same day but we constrained teachers in selecting an alternative day in a very limited time span.

- fewer missing data due to pupils did not actually do the exercise at home.

The first three points gain high importance, especially for the IAT exercise, meanwhile the latter with high probability would lead to biased data collection where missing values are related to the characteristics of the respondents[5] (MAR=missing at random).

4. DATA

4.1 Sample identification

4.1.1 Sample size: ex-ante calculation

Before conducting the experiment, a statistical power analysis was performed to verify the minimum number of students required to generate significant results. We performed a sample size estimation, using the GPower software (GPower 3.1). The software offers five types of statistical analysis and - according to our needs - we focus on the '*priori*' one in which sample size N is computed as a function of power level $1 - \beta$, significance level α, and the to-be detected population effect size.

Then we select the ANOVA (repeated measures, within-between interaction) test as we are conducting a pre-post analysis.

And finally, by following Cohen's (1988) criteria we set the expected effect size at 0.1[6] the power level at 0.8 and the alpha at 0.05.

Figure 2a reports, as an output of GPower, the minimum number of observations required with our parameters and setting. The sample size needed to detect a small effect that consists of at least 200 students (Actual power =0.8036) between treated and controls. Assuming equidistribution between treated and control optimizes data availability, the final sample consists of at least 100 treated and 100 control.

Considering this as the minimum numerosity and assuming the physiological interferences that characterise these experiments (attrition, missed take up[7], incorrect questionnaire fulfilments and the mismatch between presences in pre and post surveys) we opt for including in our experiment at least 11 class treated and 11 controls. So, we planned to include at least around 400 students

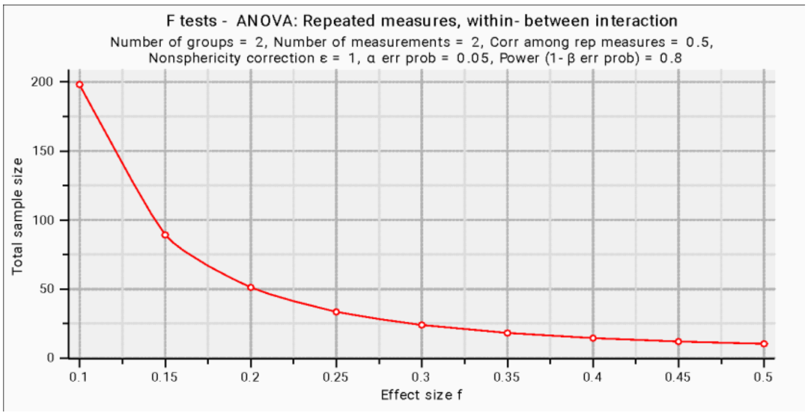
[5] It is reasonable to assume that the students who do not return the questionnaire are: the least diligent, the most careless, those with poorer family circumstances (no device), etc.

[6] Which is the size suggested for an expected small effect.

[7] As the experiment took place in classes during the official hour of lesson this condition could verify only with absences: is not so reliable to assume that a student is in class without participating. Supply treatment by classes also make impossible crossover effects.

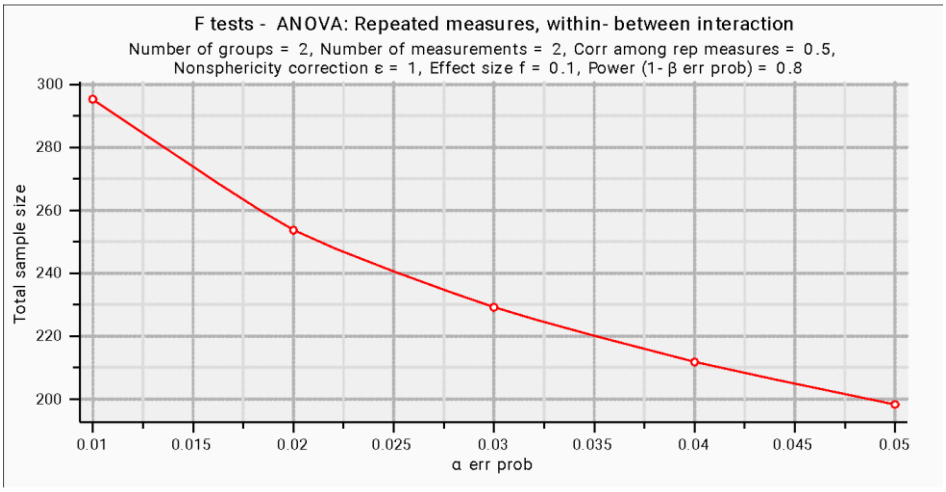
depending on class numerosity; this sample also allows for a higher level of precision: as Figure 2b shows, a sample of nearly 300 students would allow for an alpha of 0.01 instead of 0.05.

Figure 2a – Required sample size for different effect sizes



Source: Authors' computation using GPower software.

Figure 2b – Required sample size for different alpha



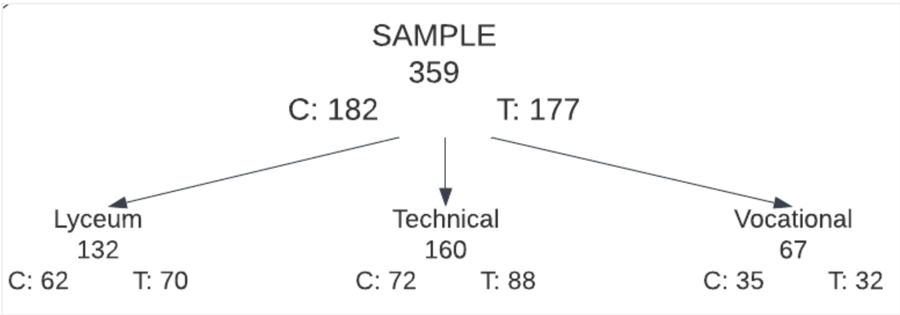
Source: Authors' computation using GPower software.

4.1.2 Sample size: final collected sample

The final sample consisted of 26 classes covering 6 schools and comprising 532 students. In the sample, 13 classes (273 students) were treated and 13 (259 students) were used as a control. We succeeded in merging 366 students (meanwhile for 98 of them we get only the first survey and for

75 only the second one).[8] Of the 366 merged data, 182 were controls and 184 were treated and among these latter 146 attends all the meetings of the treatment meanwhile 177 attended at least 2 out of 3 meetings. We decided to omit from the final sample all treated who participated in only 1 out of 3 meetings, so our final sample is composed of 359 pupils between controls and treated who had participated in at least 2 out of 3 meetings. Our final sample, represented in figure 3, includes 132 Lyceum students (36.8%), 159 technical students (44.6%) and 67 vocational students (18.7%).

Figure 3 – Sample specification



Source: Authors' computation on acquired data.

We use self-acquired primary data which were collected mainly by submitting students a questionnaire before and after the treatment. The questionnaire was developed in Qualtrics[9] after careful reflections on the most crucial information required for the evaluation of the experiment and a deep exploration of the literature. We used the latter both to extract some commonly used and scientifically validated questions and as inspiration to create queries customised to our case.

The same questions were submitted twice in the pre-survey and the post-survey with the aim of creating a panel database; the only exceptions were variables invariant over time (such as demographics and parents' information) which were asked only once in the pre-questionnaire and queries on project satisfaction which could be asked exclusively in the post and only for treated.

The questionnaire includes questions on demographics, skills, attitudes, beliefs, stereotypes, inclination toward economics, university choice intentions and drivers of the latter.

Moreover, we integrate the collected data with some information on the class composition: the number of students and the percentage of females, hours of math and economics, gender of the Math and Economics teacher, whether the class was already involved in activities aimed to diminish gender stereotypes and how much the professors seemed to be involved and believe in the project

[8] The sum of the three is 539 which is slightly higher than 532 this is since maybe some students were present both at the pre and the post surveys, but we couldn't aggregate their responses due to omitted, incorrect, or undecipherable merges codes.

[9] https://www.qualtrics.com/it/ .

(this last piece of information was gathered by submitting to the staff and those who had direct dealings with professors a short questionnaire).

In Section 4.2 we provide further insight into the main dimensions considered.

4.2 Variables definition

UNIVERSITY TRACK (or work) INTEREST

The major choice is our main outcome of interest: we ask students how likely they are to take 15 different paths after high school. Among this set of options, the first one was to go to work and the other 14 concerned different university paths. For each alternative, the answer options range from 1 (*Extremely Unlikely*) to 5 (*Extremely Likely*). Moreover, we also ask them about their certainty concerning their future (how confident they feel about what they declare in the 15 paths).

STEREOTYPES

The problem with measuring explicit stereotypes is that people may not want to reveal their thoughts (social desirability), or they are not fully aware (conscious) of their feelings.

So, we decided to measure students' stereotypes both implicitly (*Dscore*) and explicitly (*ExplicitStereotypeSchool* and *ExplicitStererreotypeFamily*).

The explicit stereotype is measured regarding the female-economics association both at school (Tomasetto, Mirisola, Galdi and Cadinu 2015) and in the household[10] environments.

To measure implicit stereotypes, we rely on the most widespread measure of the strength of stereotypical associations (De Gioannis 2022b) the Implicit Association Test (IAT) developed by social psychologists (Greenwald et al., 1998). This test, besides being widely applied by social psychologists, has also spread to economists, especially for research on gender (Carlana and Corno 2021, Carlana 2019, Tomasetto 2015) and cultural background (Corno 2018) divergences. Its strength is its ability to detect biases that operate at the unconscious level (or that respondents do not want to disclose) by measuring the reaction rate. The reaction time is measured by submitting participants to a categorization task in which they are asked to place as faster as possible on the left or on the right side stimuli belonging to four categories. In our case, as we are interested in the association between gender and fields, the categories are Masculine names (target A), Feminine names (target B), Economics and finance subjects (attribute A) and Humanistic subjects (attribute B).

The higher (or lower) rate of target and attribute placement may reveal how strongly an individual associates subject domains to gender and thus the presence of stereotypes.

[10] Items were inspired by the ISTAT (2011) survey on discrimination by gender.

SUBJECTIVE TASK-VALUE

Within this dimension, being multidimensional, we used a wide and diverse range of indicators. We used the reduced scale (adapted to the financial-economic domain) of Good *et al* (2012) for the Sense of belonging containing the subscales acceptance (*SenseofbelongingIndex1*) and membership (*SenseofbelongingIndex2*).

As intrinsic values, we use two variables aimed at assessing how much students like math (*Like_maths)* and/or economics (*Like_economics*).

We also ask how much they consider Economics, Maths, Statistics and Finance important for their personal fulfilment, their future academic achievement and professional fulfilment as an indicator of utility (other goals) in the following tables these variables correspond namely to *ImpoEconomyforLife* , *ImpoEconomyforAcademic* and *ImpoEconomyforProfession*.

We also included possible costs (*cost*) that individuals might perceive with some questions from PISA (concerning that math lessons could be difficult, nervousness for solving math assignments or problems, feeling that they cannot solve a math problem and fear of getting low math grades).

EXPECTATIONS OF SUCCESS

In our model, expectations of success are based on two components: the observed ability and the own prejudice about their abilities. For the first (*MeasuredAbility*), we simply used grade point average in maths and – if available – economics. We determine the latter (*AbilitySelfConcept*) with a set of questions from PISA 2012 in which students report if they are good at math, if they get good grades or feel talented in math like "In my maths class, I understand even the most difficult work". We also integrate with the overconfidence (*Overconfidence_maths* and *Overconfidence_economics*) variable which is computed by comparing their observed ability and their declared one. According to the literature, this second component based on personal perception is a strong determinant of course enrollment and occupational aspirations choice (Bandura 1997, Bandura et al. 2001).

ENVIRONMENTAL FACTORS

In addition to class fixed effects with information on peers and professors, we collect a set of family background variables.

Family background: Family factors play a fundamental role in choices and abilities, even more so than the school itself (Bottazzi e Lusardi 2020). For both parents, we collected: their place of birth, level and field of study. We also collected information on the care and time both parents devote to the student. The latter includes items such as the time spent by parents having dinner with their children, asking them how things are going at school, shows interest in their children's future and their feeling. As income proxies, we collect the work they do, their contract type, home assets (together with study assets) and the number of books at home.

The questionnaire also includes an exhaustive battery of questions about what the students feel is more important in their choice (all environmental factors, abilities, interests, works, etc… are included). The summary statistics of student and family characteristics are reported in Table 1.

Student characteristics: Gender, age, whether they have ever repeated a grade or more, place of birth and years in the Emilia Romagna region.

4.3 Descriptive statistics

Table 1 – Summary statistics for the main covariates

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Student observable characteristics					
Female	359	0.67	0.47	0	1
age	359	19.19	0.57	18	22
YearsInEmiliaRomagna	359	18.02	3.26	2	22
RepetedClass	358	0.19	0.39	0	1
North	359	0.86	0.35	0	1
Family background					
Mother_North	359	0.50	0.50	0	1
Father_North	359	0.49	0.50	0	1
Mother_North	359	0.50	0.50	0	1
Father_North	359	0.49	0.50	0	1
Mother_Elementarymiddle	359	0.3	0.46	0	1
Father_Elementarymiddle	359	0.41	0.49	0	1
Mother_Highschool	359	0.43	0.50	0	1
Father_Highschool	359	0.38	0.48	0	1
[1] *Mother_Lyceumh*	359	0.09	0.28	0	1
Father_Lyceumh	359	0.08	0.27	0	1
Mother_Lyceums	359	0.02	0.14	0	1
Father_Lyceums	359	0.03	0.16	0	1

Mother_Technical	359	0.21	0.41	0	1
Father_Technical	359	0.19	0.39	0	1
Mother_Vocational	359	0.08	0.26	0	1
Father_Vocational	359	0.07	0.25	0	1
Mother_Universitymore	359	0.19	0.39	0	1
Father_Universitymore	359	0.12	0.33	0	1
[2] *Mother_Eco*	359	0.01	0.12	0	1
Mother_Stem	359	0.01	0.12	0	1
Mother_Hum	359	0.10	0.3	0	1
Father_Eco	359	0.02	0.13	0	1
Father_Stem	359	0.03	0.16	0	1
Father_Hum	359	0.04	0.20	0	1
Mother_Notemployed	359	0.03	0.17	0	1
Mother_Housewife	359	0.15	0.36	0	1
Father_Notemployed	359	0.05	0.22	0	1
Mother_Empl1 [3]	359	0.26	0.44	0	1
Father_Empl1 [3]	359	0.36	0.48	0	1
Mother_Empl2	359	0.42	0.49	0	1
Father_Empl2	359	0.23	0.42	0	1
Mother_Empl3	359	0.04	0.19	0	1
Father_Empl3	359	0.12	0.32	0	1
Mother_Empl4	359	0.07	0.25	0	1
Father_Empl4	359	0.18	0.39	0	1
MumCare	358	4.09	0.95	1	5
DadCare	358	3.55	1.2	1	5
WellnessHome[4]	359	5.4	0.88	0	6
WellnessStudy[5]	359	4.73	.67	0	5
Bookathome0to10	359	0.11	0.31	0	1
Bookathome11to25	359	0.14	0.34	0	1
Bookathome26to100	359	0.31	0.46	0	1

Bookathome100more	359	0.35	0.48	0	1
School environment					
Lyceum	359	0.37	0.48	0	1
Technical	359	0.45	0.5	0	1
Vocational	359	0.19	0.39	0	1
classSize	359	21.01	3.48	13	26
ShareFemale	359	59.78	25.67	11.11	94.12
FemaleTeacher	359	0.77	0.42	0	1
TeacherBeliefInTheProject	359	7.06	1.69	4.67	10
TeacherMotivateStudent	359	6.63	2.21	3	10
TeacherEngagement	359	6.83	2.01	3	10
TimeSpan	359	11.18	4.59	-2	18
Hmath	359	16.13	1.6	12	17
Heco	359	20.28	12.49	0	33

Note: The sum of the percentages of the complementary variables (such as education and occupation of parents, book at home, etc..) could be lower than 100% as students have the possibility to answer "I don't know".
[1] High school specification: the sum of mum (or dad) specification percentages returns the mum (or dad) percentages with a high school diploma.
[2] University specification: most relevant group selection (dichotomies that only consider degrees in the humanities, STEM or economics).
[3] Work specification: for both parents are four dummies variables which indicate a labour status which increases with the number (further details on categories are included in the table A3 in the annexes.
[4]WellnessHome: continuous variable ranging from 0 to 6 indicating the possession of some basic "home assets" as dishwasher, fridge, oven, etc..
[5]WellnessStudy: continuous variable ranging from 0 to 5 indicating the possession of some "study assets" like PC, dictionary, internet connection, room and a table intended for study.

The sample consists of 67% females, this partly dictated by decision-making choices related to the nature of the intervention that led us to exclude classes with few (or no) girls.

Students were mainly born in Northern Italy (85%) but, for their parents, this percentage falls below 50 per cent: around 30% came from Southern or Central Italy and the remaining 20 per cent emigrated from abroad.

Regarding the family background: parents' educational qualification is mainly concentrated in middle school diploma (30% of mothers and 40% of fathers) and high school diploma (40% of mothers and 37% of fathers). Confirming that, even in the previous generation, mothers are the more educated ones.

The same is found for the university degree which is possessed by 19% of mothers compared to 12% of fathers. Gender distribution across fields of study shows that, even for the previous generation, there were gender imbalances among the various areas (women engaged in highly mathematical fields were only $\frac{1}{3}$ of those engaged in humanities fields while for fathers the same ratio is $\frac{3}{2}$). The percentage of unemployed mothers exceeds 18% compared to only 5% for fathers, moreover fathers' not working status is mainly due to retirement (and associated with a pension) while that of mothers is related to being housewife. Within employed parents, households mothers are concentrated at the lowest levels of employment and the percentage of fathers in the highest one is about three times that of mothers. Finally, there are parental differences also in the time and attention devoted to children: care variables show that, taking fathers as a basis, mothers have 15% more care (time spent together, talking, worrying about students' feelings and their future) for their children.

This family background snapshot - in line with expectations, and the Italian context - is a fertile ground for implicit stereotypes: students are born and grow up in a context that becomes 'normal' for them but which, as we can see, is far from being gender-balanced. It is not a negligible influence as the home environment has a strong influence on students' decisions, more than the school environment (Bottazzi Lusuardi 2020).

5. RESULTS

5.1 Gender differences in major choice determinants

Table 2 allows analysing of the differences by gender in the variables that reveal stereotypes of field choice and degree of confidence in the different fields. As can be seen, by the observed average and statistically significant differences among them, female students are more likely to have higher implicit stereotypes[11] as revealed by IAT (Dscore) while, at the start of the experiment, they show lower explicit stereotypes revealed by two variables that refer to the school and to the family environment. However, there does not appear to be a statistically significant difference by gender in math or economics enjoyment nor in the ability self-concept. On the other hand, male students assign higher importance to economics in life and have lower psychological costs (fear to get low grades in math, nervousness and|or helplessness and|or anxiety) in addressing subjects with higher math content. Moreover, males tend to be more overconfident than female students regarding math and economics.

[11] This is in line with previous research with a similar setting [De Gioannis 2022a].

Table 2 – Gender comparison on major choices determinants AT T=0

	(1) FEMALE		(2) MALE		(3) T-TEST ($\bar{x}_m - \bar{x}_f$)	
	mean	sd	mean	sd	b	t
Dscore	0.41	0.40	0.17	0.39	-0.24***	(-5.39)
ExplicitStereotypeSchool	2.21	1.17	2.80	1.27	0.59***	(4.20)
ExplicitStererreotypeFamily	1.37	0.69	2.55	1.26	1.17***	(9.43)
MeasuredAbility	7.32	1.20	6.73	1.16	-0.59***	(-4.49)
AbilitySelfConcept	2.39	0.75	2.45	0.74	0.06	(0.70)
Like_maths	3.29	1.67	3.47	1.56	0.17	(0.95)
Like_economics	3.46	1.61	3.82	1.54	0.36	(1.73)
ImpoEconomyforLife[1a]	2.94	1.34	3.24	1.19	0.30*	(2.16)
ImpoEconomyforAccademy[1b]	2.85	1.35	2.94	1.18	0.09	(0.61)
ImpoEconomyforProfession[1c]	3.17	1.40	3.49	1.19	0.32*	(2.22)
Cost	2.62	0.77	2.25	0.68	-0.37***	(-4.58)
SenseofbelongingIndex1 [2a]	4.59	1.23	5.08	1.12	0.49***	(3.75)
SenseofbelongingIndex2 [2b]	3.82	1.75	4.45	1.67	0.63**	(3.29)
Overconfidence_maths	0.19	0.39	0.31	0.46	0.12*	(2.36)
Overconfidence_economics	0.14	0.34	0.26	0.44	0.13**	(2.76)
N	242		117		359	

Notes: Because these variables change over time and may be affected by treatment, we report values at the first survey (T_0) before the start of the project).

[1]*This set of variables corresponds to students' rate from 1 (Not important) to 5 (Very important) at the following question:*

How important do you think economics, mathematics, statistics, finance are for your future prospects? [1a] *future personal fulfilment (of yourself as a person and in everyday life);* [1b]*future academic achievement (at university);* [1c]*future professional fulfilment (working career)*

[2]*Composite indexes are based on a set of continuous variables ranging from 1 "low sense of belonging to the financial-economic community" to 8 "high sense of belonging to financial-economic community"* [2a] *subscale acceptance;* [2b]*subscale membership.*

Implicit and explicit stereotypes

Table 3 shows the correlation between the implicit stereotype (Dscore, column 1), the explicit ones (in the family (2) and school environment (3)) and other choices drivers together with the enrolment propensity for finance, marketing, stem and humanistic fields.

As we can see from the first two rows of column 1 in both panels of the table, regardless of gender, the implicit and explicit stereotypes do not correlate. This result is common knowledge in social psychology: in fact, the IAT test is used exactly to gather more accurate information through answers which are not biased by scarce awareness and|or social desirability. In fact, as shown by panel A (female) in columns (2) and (3) the correlation between choices determinants and preferences is basically absent, meanwhile, the IAT results exhibit a very strong correlation with most of them. Those results are in line with Carlana and Corno (2021) who found a positive correlation between the IAT and the choice of the field, but they didn't find the same between IAT and the explicit stereotypes and neither between the explicit and the math choice.

On the other hand, males – in panel B – show a correlation between implicit stereotypes, some determinants and enrolment intention. Our thought is that they actually were more transparent in (or conscious of) their explicit revelation. Recalling the results of table 2: meanwhile, females had a high implicit stereotype and a low explicit one, males returned more constant averages in the two types of stereotypes. Moreover, in table 3 panel B shows a higher male correlation (which is also nearly significant: p-value not reported = 0.11) between the explicit and the implicit (school) stereotypes.

Moreover, as expected, it is very interesting to state that the stereotype interacts with the main dimensions and outcomes in opposite ways depending on gender (positive relationship for males and negative for females). This is because suffering from stereotypes means associating females with humanistic subjects and males with economical and mathematical ones, so stereotypes affect self-confidence and beliefs in the opposite ways depending on one's gender. It derives, as can be seen from Panel A of Table 3, that a girl who suffers from high stereotypes rejects the numerical field: having low self-confidence, feeling it is less important, stating that likes it less, avoiding enrolment in business degrees and preferring humanities degrees. On the other hand, males (panel B) who possess stereotypes tend to enrol more in economics fields and perceive lower difficulties and they consider economics more important.

Table 3 – CORRELATION between implicit and explicit stereotypes and major choice determinants

	(1)	*(2)*	*(3)*
	Dscore	*EXP[ ]SCHOOL*	*EXP[ ]FAMILY*
PANEL A - FEMALE			
ExplicitStereotypeSchool	0.08		
ExplicitSterereotypeFamily	-0.03		

MeasuredAbility	-0.06	-0.18**	-0.10
AbilitySelfConcept	-0.07	-0.12	0.01
Liike_maths	-0.15*	-0.07	0.10
Like_economics	-0.25**	-0.12	0.02
ImpoEconomyforlife	-0.20**	-0.07	0.06
ImpoEconomyforAcademy	-0.26***	-0.07	0.08
ImpoEconomyforProf.	-0.29***	-0.06	0.16*
Cost	-0.02	0.11	0.02
SenseofbelongingIndex1	-0.20**	-0.03	0.12
SenseofbelongingIndex2	-0.15*	-0.02	-0.01
Overconfidence_maths	-0.10	-0.01	0.03
Overconfidence_econ.	-0.22***	-0.07	0.07
Finance[1]	-0.25***	-0.11	-0.01
Marketing[1]	-0.26***	-0.14*	-0.09
STEM[1]	-0.14*	-0.05	0.11
Humanistic[1]	0.14*	0.11	-0.08
PANEL B - MALE			
ExplicitStereotypeSchool	0.11		
ExplicitStererreotypeFamily	-0.08		
MeasuredAbility	0.02	0.12	-0.11
AbilitySelfConcept	0.18	0.17	-0.12
Liike_maths	-0.01	0.10	0.03
Like_economics	0.09	0.19	0.00
ImpoEconomyforlife	0.20*	0.19*	0.16
ImpoEconomyforAcademy	0.22*	0.11	-0.15
ImpoEconomyforProf.	0.18*	0.18*	0.07
Cost	-0.19*	-0.16	0.14
SenseofbelongingIndex1	0.13	0.29**	-0.01
SenseofbelongingIndex2	0.08	0.31***	0.12
Overconfidence_maths	-0.04	0.11	-0.03
Overconfidence_econ.	0.08	0.19*	0.11

Finance[1]	0.31^{***}	0.35^{***}	0.06
Marketing[1]	0.29^{**}	0.31^{***}	0.09
STEM[1]	0.17	0.01	0.01
Humanistic[1]	-0.04	-0.00	-0.13

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Notes:
[1] *likelihood of enrolling in this specific field*

The expectation of success: ability and ability self-concept

Table 4 – Gender differences in ability

Gender	Overall[1]	Maths	History	Italian	Economics	English	Philosophy
Male	7.18	6.67	7.10	6.98	6.69	7.08	7.04
Female	7.61	7.32	7.43	7.46	7.29	7.50	7.33

[1] *Overall is not the average of the grades shown, but the overall mark on the school report.*

Despite indicators in table 4 show that females outperform men in each dimension (economics and mathematics included), female overconfidence in math and economics is scarce and extremely lower than the male one ($Pr_{(|T| > |t|)} < 0.01$) condition that does not occur for subjects such as English, history, and Italian where both genders have similar self-confidence values. Moreover, females tend to disagree with self-praising and self-confident phrases such as *"In my maths class, I understand even the most difficult work"* ($Pr_{(|T| > |t|)} < 0.05$) or "*I learn mathematic quickly"* ($Pr_{(|T| > |t|)} < 0.01$).

Subjective task value

As for reported literature, this dimension is composed of reasons that motivate (engage) individuals. Falls among these intrinsic motivations is the enjoyment of the subject and the extrinsic one which consists of the perceived utility in studying it.

Table 2, shows that - even though girls have lower scores - there are no significant gender differences in how much students like math (*Like_maths*) and economics (*Like_economics*). The same condition occurs in the answers to the question "I have always believed that mathematics is one of my favourite subjects" inspired by PISA questionnaire. Regarding extrinsic motivation, there are significant differences in favour of males on the thought that economics is important for life and profession (respectively *ImpoEconomyforLife* and *ImpoEconomyforProfession* in table 2) and no statistical difference in the perceived utility of economics for the academic life although, again, females show a lower score.

Another dimension falling under the subjective task value is the composite index of the sense of belonging which is significatively higher in males for both sub-dimensions: acceptance (*SenseofbelongingIndex1*) and, even more, membership (*SenseofbelongingIndex2*).

Last but not least, costs are another important component which affects the subjective task values in a negative way: the higher the fear of the subject and the perception of inadequacy, the lower the motivation toward that pathway. Once again, from table 2, we see females in a disadvantaged condition with significative higher costs. As the composite index (*cost*) all the cost sub-variables show high gender differences. The greatest ones are related to anxiety: the rate of females in the statement *"I am very nervous when I must do a math assignment"* and *"I am very nervous when I must solve a maths problem"* are respectively 19.2% and 22% higher than the males one both with a $Pr_{(|T| > |t|)} = 0.000$.

5.2 Relationship between choice determinants and enrolment intention

The dimensions introduced in table 2 and analysed so far have shown a strong gender disparity, even though the observed abilities are equal if not even favourable to females. The purpose of Table 5 is to investigate how these determinants influence the propension to enrol in a finance (column 1), marketing (column 2), STEM (column 3) or humanities (column 4) degree.

Table 5 – Correlation between choices determinants and major choice preferences

	(1) FINANCE	(2) MARKETING	(3) STEM	(4) HUMANISTIC
PANEL A - FEMALE				
MeasuredAbility	0.25***	0.23***	0.29***	-0.08
AbilitySelfConcept	0.25***	0.22***	0.35***	-0.21**
Liike_maths	0.28***	0.27***	0.41***	-0.23***
Like_economics	0.67***	0.62***	0.35***	-0.29***
ImpoEconomyforlife	0.43***	0.37***	0.31***	-0.23***
ImpoEconomyforAcademy	0.52***	0.53***	0.38***	-0.14*
ImpoEconomyforProfession	0.51***	0.55***	0.39***	-0.24***
Cost	-0.12	-0.07	-0.28***	0.09
SenseofbelongingIndex1	0.41***	0.41***	0.31***	-0.13*
SenseofbelongingIndex2	0.54***	0.50***	0.36***	-0.09

Overconfidence_maths	0.18**	0.12	0.17**	-0.11
Overconfidence_economics	0.35***	0.31***	0.18**	-0.03
PANEL B - MALE				
MeasuredAbility	0.20*	0.18	0.18	-0.10
AbilitySelfConcept	0.22*	0.20*	0.39***	-0.05
Liike_maths	0.15	0.13	0.47***	0.02
Like_economics	0.63***	0.53***	0.31**	0.18
ImpoEconomyforlife	0.40***	0.36***	0.16	-0.11
ImpoEconomyforAcademy	0.55***	0.54***	0.40***	-0.03
ImpoEconomyforProfession	0.46***	0.48***	0.31***	-0.14
Cost	-0.14	-0.15	-0.19*	0.03
SenseofbelongingIndex1	0.41***	0.36***	0.12	-0.04
SenseofbelongingIndex2	0.42***	0.35***	-0.01	-0.08
Overconfidence_maths	0.02	0.01	0.32***	0.06
Overconfidence_economics	0.39***	0.38***	0.33***	0.12

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: stereotypes are omitted because the relationship is already reported in table 3.

For both genders, there is evidence that the dimension selected as choice determinants have significant relationships with the propensity to enrol in different university courses. In fact, consistently with the literature analysed, enrolment decision-making does not appear to be simply a matter of ability and skill. Indeed, the correlation between measured ability and choices shows the lowest values for both sexes becoming of low or zero significance for the male subsample.

For females - Panel A - the stronger relationship with the chosen path is given by how much they like the subject of economics (correlation with the likelihood of enrolment in economics highly significant and greater than 0.6) or maths (highest correlation between all determinants with the STEM enrolment likelihood). Females who are inclined to enrol in economics degrees are also those who consider economics to be more important for their academic and working future and life in general and the ones who feel to belong to the financial-economics community (SenseofbelongingIndexes). Overconfidence in economic capabilities is also important for economic enrolment and has a correlation with this latter greater than the actual ability itself. On the other hand, female overconfidence in maths has a lower correlation; which disappears for the marketing degree (which has a lower quantitative content) but is still consistent for the finance and STEM degree.

Surprisingly, despite the expected sign, the costs -which for girls are an emerging problem- do not correlate significantly with the propensity of enrolment in economics. It is possible to explain this evidence by considering that the cost index was built on variables concerning exclusively the math dimension. In fact, we can see that for the propensity to enrol in STEM degrees this component comes to be as important and significant as the ability itself.

Furthermore, in the female sub-sample, the selected determinants correlate satisfactorily well also with the probability of enrolment in humanities degrees. The appreciation of economic-mathematical subjects, the perception of one's own abilities and the importance given to economic subjects: all have a negative and highly significant relationship with enrolment in humanities.

For males (panel B), the math component seems to be less relevant in the intention of enrolment in economics as there is no correlation between it and both the overconfidence and appreciation of math. On the other hand, these latter are very important for the propensity to enrol in STEM. The relationship between the remaining determinants and enrolment in economics does not diverge from the results seen in the female subsample. Conversely, the relationship between determinants and the propensity toward a humanistic path (column 4) strongly deviates from female evidence.

None of the dimensions analysed seems to be linked to the male choices in the humanistic field. Even though determinants are purpose-built to study enrolments in economics degree courses, the complete absence of a relationship with the humanistic degree is an unexpected and relevant fact: given the complementary nature of the two types of fields of study, it is plausible to assume the presence of an inverse relationship.

From this table, it is not possible to explain what drives young males toward the humanities, but it does not seem to be related to ability, nor of appreciation or usefulness.

5.3 Treatment effects

Change in choice of the majors

Figure 4a shows, for treated and controls (vertical axes), the *shift* in the probability of enrolment in different fields (horizontal axes). The latter is derived by the pre-post difference of the average score in the respective choice's intention.

Males treated -compared to control- have a lower increase of their certainty with respect to their future choices (green pipe): from the beginning males show high confidence about their decision, but it seems that, by participating in the project, they have begun to re-examine themselves. the most interesting result in the treated males is precisely the decrease in the propensity to work after high school (whereas the controls show an increase in this variable over time). Finally, they decrease the desire to enrol in finance and increase the propensity toward humanities and marketing degrees. Turning now to female students in figure 4b treated girls show greater awareness (green pipe) than female students in the control group. Almost no differences in the propensity to go to work can be observed in the two groups it slightly decreases in both. The results for the fields of study are interesting and positive: in fact, two cross trends in treated and control groups can be noted. The first concerns economics, where there is a switch between finance (which is a highly quantitative course) and marketing controls; females in the control group seem to be more oriented toward marketing and treated toward finance. The second regards the STEM and humanities degrees: in fact, while there is no substantial difference in the propensity for humanities degrees, the propensity for STEM degrees is much stronger for the girls treated and shows an increase that even surpasses the humanities degrees' likelihood.

Figure 4a – Temporal shift in choice: from T_0 to T_1

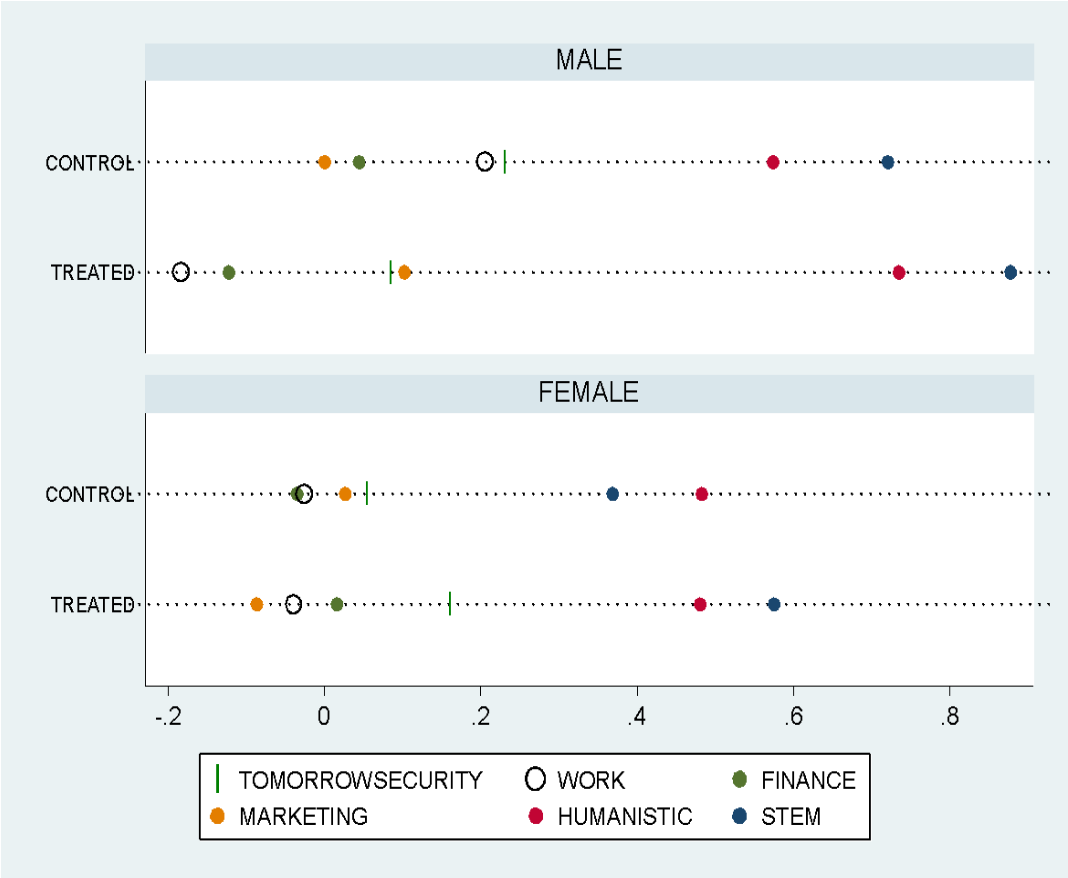


Figure 4b offers further insight by splitting the analysis by school type (Lyceum, Technical, Professional). It is discernible how shifting in choices are heterogeneous according to the typology of the institute: Lyceum students show more restrained displacements; this could be due to the fact that perhaps in high school they already have more solidly rooted ideas: perhaps more confidence, more organisation, an approach that leads them to make decisions in advance, etc…Students from technical and professional high schools, indeed, seem to have a more malleable mind: both treated and controls show more pronounced changes in their choices, suggesting that the decision-making process is still open and evolving.

Treated male students in technical institutes highly increase their propensity towards humanities fields if compared with the counterfactuals and females increase the one towards STEM with the same intensity as humanities: situations which do not occur for their counterfactual. In technical institutes both males and females treated groups show a decrease in their work intentions, especially if compared with the control groups.

Figure 4b – Temporal shift in choice: from T_0 to T_1 - SchoolType details

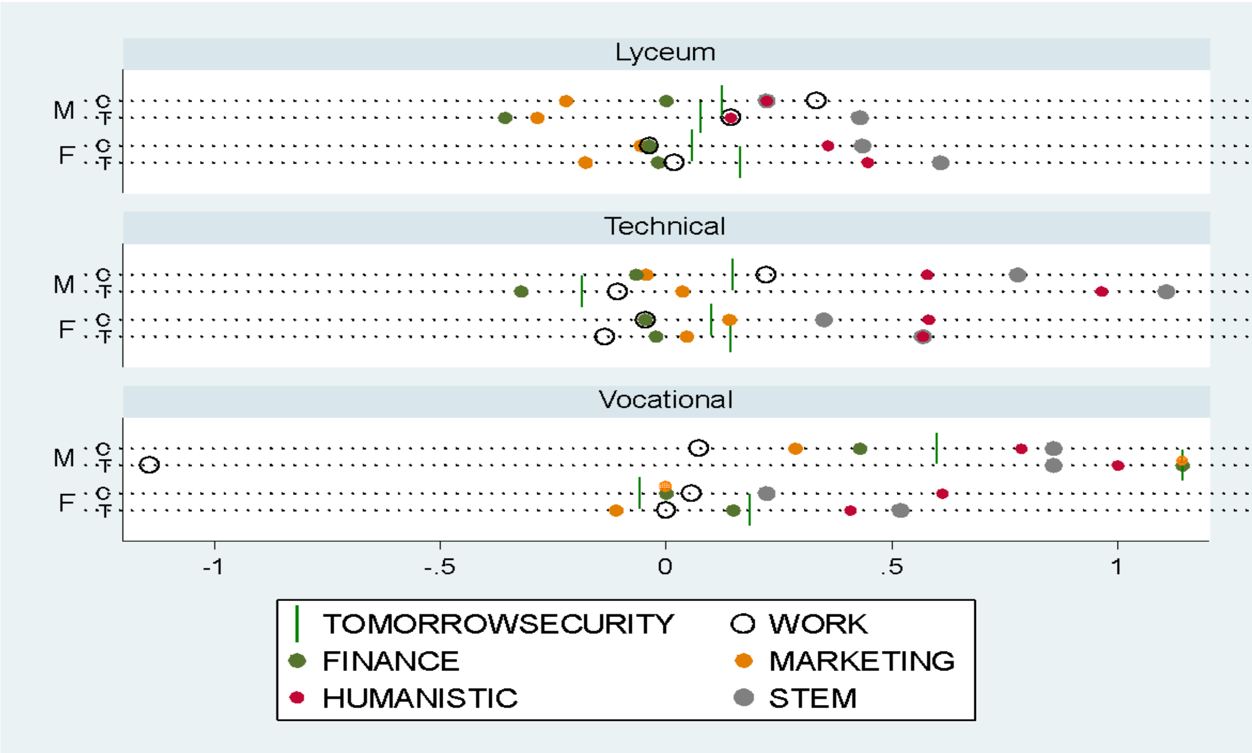


Figure 4b offers further insight by splitting the analysis by school type (Lyceum, Technical, Professional). It is discernible how shifting in choices are heterogeneous according to the typology of the institute: Lyceum students show more restrained displacements; this could be due to the fact that perhaps in high school they already have more solidly rooted ideas: perhaps more confidence, more organisation, an approach that leads them to make decisions in advance, etc…Students from technical and professional high schools, indeed, seem to have a more malleable mind: both treated and controls show more pronounced changes in their choices, suggesting that the decision-making process is still open and evolving.

Treated male students in technical institutes highly increase their propensity towards humanities fields if compared with the counterfactuals and females increase the one towards STEM with the same intensity as humanities: situations which do not occur for their counterfactual. In technical

institutes both males and females treated groups show a decrease in their work intentions, especially if compared with the control groups.

The first peculiarity of students in vocational institutes is that both genders, when treated, have a higher increase in their confidence in their future with respect to counterfactual. Another remarkable change is in male orientation toward work: meanwhile the counterfactual remains stable around zero (no change in their tendency) the one of treated suffers a huge decrease (-1.14, which is nearly to 30% of the maximum possible shift from "extremely likely" to "extremely unlikely"). Treated males also show a great increase in their propensity towards humanities, but the biggest one is towards economics (both marketing and finance). The evidence on males makes us assume that they perceived this external intervention as a strong stimulus that got them involved and made them consider other options besides work. Girls attending vocational high schools, on the other hand, show the same positive trend as the sample as a whole, but with a major emphasis.

Figure 5 – Insight in Work likelihood: treated and control in T_0 and T_1 - schoolType details

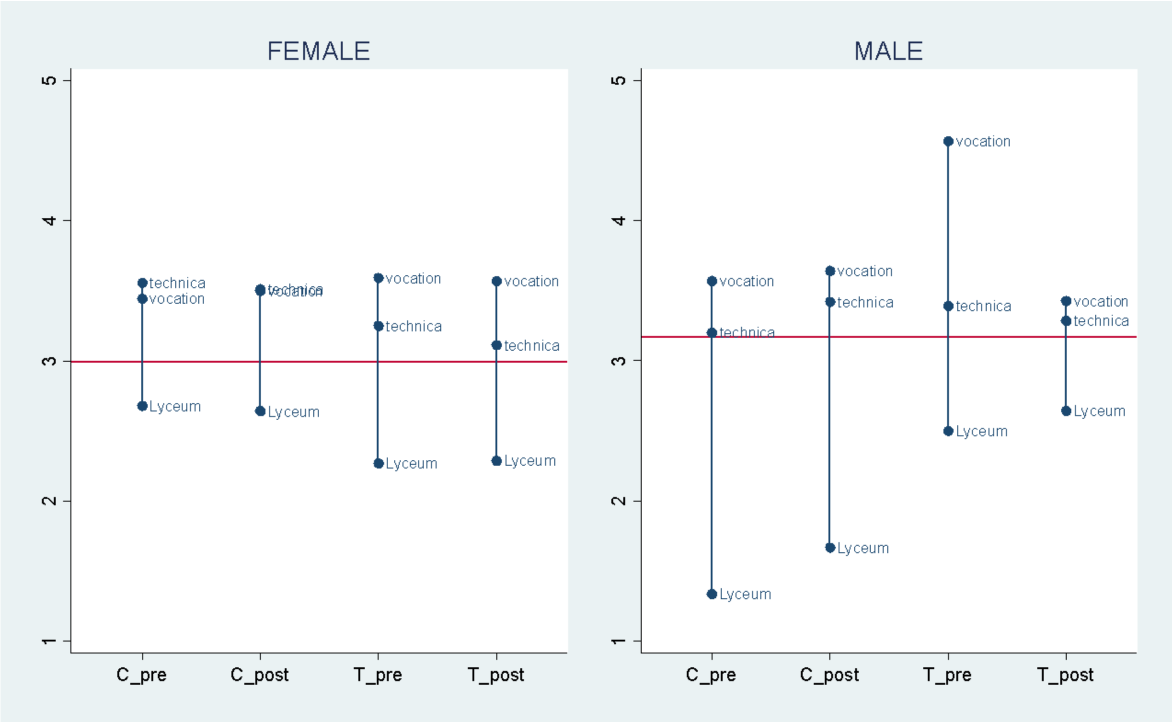


Figure 5 reports a focus on the propensity to work by the typology of the institute. On the left, we have the female subsample and on the right the male one. On the horizontal axes there is the control group in the survey at T0 (C_pre) and at T1 (C_post) and the treated one before (T_pre) and after treatment (T_post), while on the vertical axis it shows the propensity to go to work where 1 corresponds to "extremely unlikely" and 5 "extremely likely".

As expected, the lyceum is always placed at the bottom of each subsample and vocational schools are almost always at the top. On average (red line) males are slightly more likely to go to work than females (except for the lyceum control group where a very low male propensity to work can be seen). Females who participated in the project did not experience significant changes in their propensity to work, but we can see a slight decrease for the students of technical and a substantial stability in the ones attending vocational high schools while female students in the control groups attending vocational high school show a modest increase in the propensity to work.

The effects on males (shown in the right side of the figure) are more noticeable, at first treated in vocational high schools show a big drop in the propensity to work with regards to their control counterparts whose propensity to work remains almost stable. Also, the treated group attending technical high schools shows a decrease in the probability of working while their control counterpart shows an increase in the propensity to work. Finally, there is a small increase in the treated male students attending Lyceum, but this is still lower than that shown by the control group.

Satisfaction/engagement

We also ask directly to students in the treated group their subjective thoughts about the project and every single session.

A positive common opinion emerged, especially for girls who appear to be the most satisfied about the project and those who on average rate the meetings the highest.

Very few participants (less than 7%) declare that they were unhappy to get involved in the project, and this percentage falls below 4% for the female population.

For each of the three sessions (board game, short film and role model) students are called to rate from one to five (the higher the score the more positive the answer) the following four questions:

1. Do you like it?
2. Do you think that the activity you have just done has changed your beliefs and perceptions? (belief1)
3. Do you think that the activity you have just done will influence your choice of university? (belief2)
4. Do you think the activity has changed your view of yourself in the world of work? (belief3)
5. In general, how useful do you consider it?

Table 6 reports average scores by gender: generally, positive opinions emerged, especially for girls who are the ones who enjoyed the activities most and found them more useful.

The activities were highly appreciated (like), especially the board game, which is very close to the maximum score of 5. Of the five-dimension observed, the second one which -for all the activities- gets the highest score is the perceived utility.

The role model meeting got the highest scores in all dimensions concerning perceived usefulness and influence. The t-test comparison reveals significant gender differences in the role model utility (p-value <0.05) and appreciation, but this result seems to be driven by a high female rate rather than a low male engagement.

In fact, despite the role model speech being the only activity targeting uniquely females, we get pleasantly impressed by discovering that males not only like it but also feel it was the most useful one and the one that most shape their perceptions, future choices and self-view at work.

Table 6 – Satisfaction

	Like	Belief1	Belief2	Belief3	Utility
Free to Choose					
Male	4.29	2.44	2.00	2.07	2.35
Female	4.36	2.54	2.02	2.41	2.60
Short on Work					
Male	3.39	2.71	2.00	2.22	2.76
Female	3.74	2.68	2.08	2.51	2.92
Role Model					
Male	3.42	2.78	2.53	2.40	2.89
Female	3.79	2.68	2.52	2.60	3.21

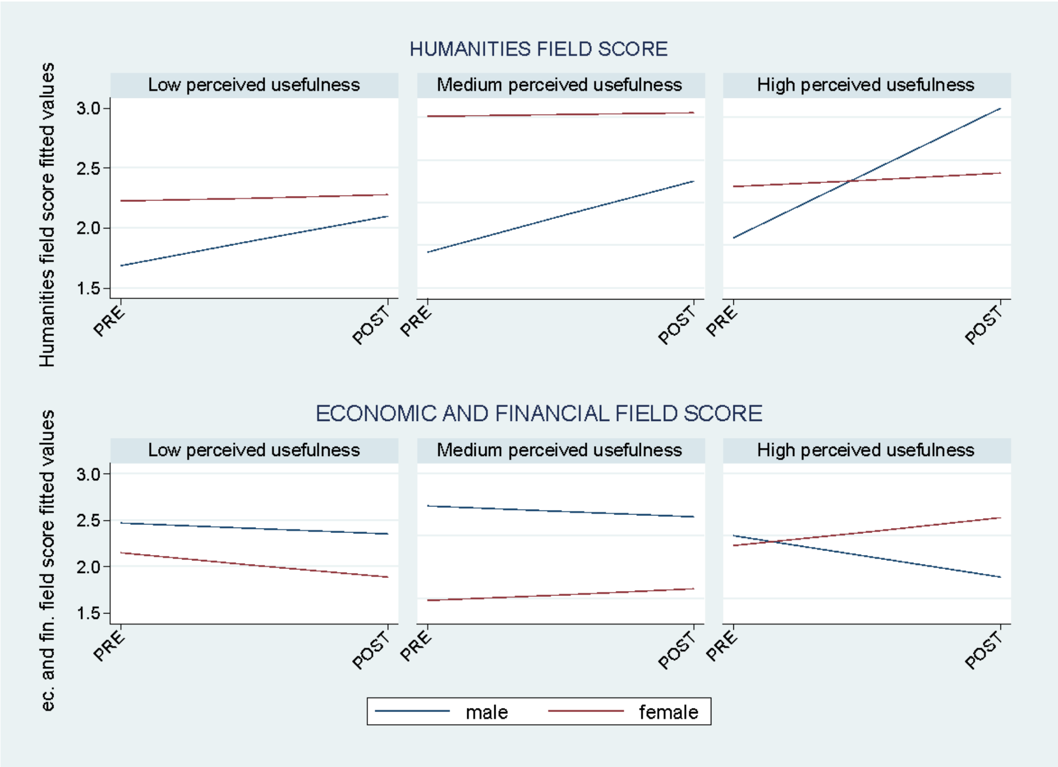
Figure 6 illustrates by gender and survey time (*pre* and *post*) the fitted values of the propensity to enrol in a humanities field (above) and in the financial economics one (below). The analysis is conducted on treated which are divided into three subgroups according to personal perception of the project's utility.

At first, it is important to point out that the starting conditions (pre) were not homogeneous in terms of perceived usefulness: girls who found the project most helpful had a greater initial propensity towards financial economics as well as a lower one towards humanities and, on the other side, males show exactly the opposite. Slopes in red and blue lines show changes over time.

The upper side of the tables suggests that female students who participated in the project maintained their inclination toward humanistic subjects basically unchanged, while males in all groups report a higher propensity towards humanities degrees with the highest increase for those who found the project more useful.

The lower side of the figure shows that both males and females who perceive the project as not very useful slightly decrease their interest in a finance degree, while there is a positive reversal for females who find the project moderately useful and, finally, trends became more pronounced for those who found the project very useful.

Figure 6 - Satisfaction and outcome



Our findings would seem to show that the students who believe the most in the project are also the ones who show the higher starting conditions toward the "*not expected field*" and have the greatest changes.

Among students who perceive the project as very useful, there has been a real reversal: male propensity towards humanities exceeds female propensity (upper part of figure 6) and the female tendency toward finance outnumbers the males' one (lower part of figure 6). The idea -coming from the starting condition – is that this particular sub-group is exactly the one formed by male and female students who would like to place themselves outside the *"stereotype box"*, feel more represented by the project, more involved and ultimately freer in their choices. So, we might see the switch in high perceived usefulness as a reallocation, closer to an optimal point, where there is a better match between choice and personal abilities and interests.

6. CONCLUSIONS

In this paper, we have analyzed the determinants of college major choices with a gender perspective, with the aim of investigating to what extent students in their last year of high school are entrenched by gender stereotypes in their choice of whether to carry on with their studies or to look for a job or, having chosen to pursue a tertiary education, how their college major choices are affected by a number of different factors including explicit and implicit gender stereotypes.

For these purposes, a field experiment has been carried out in a set of high schools in Italy, a country where the gender gap in employment at the disadvantage of women is higher than the EU-27 average and with a high level of employment segregation by gender. The experiment has been conducted in two districts of the Emilia Romagna region (Modena and Reggio Emilia), where female labour supply is sensibly higher than on average in Italy, but - where the labour market shows signs of employment segregation and inequalities in time allocation by gender.

The family background snapshot of our sample matches this scenario. The majority of mothers have undertaken humanistic studies, spend more time with their children, have lower-paid jobs or are housewives; on the contrary, the majority of fathers have scientific backgrounds and better-paid jobs (even if they hold on average a lower-level of education). Within the high schools of our sample, the composition of the classes is unbalanced by gender, with a large concentration of females in less quantitative courses, and vice versa for males. In line with the literature, this evidence has non-negligible implications for our experiment: environmental factors (such as school and family context) play a determining role in self-perception and choices. Therefore, students cannot be "neutral" in their choices, having already distorted beliefs that have been forming since their birth within their family and during their studies.

Our descriptive statistics show that the choice of the type of high school (lyceum, technical or vocational) is not the only bias that keeps girls away from subjects such as economics and mathematics: female students consider economics to be less important for their daily life and profession, have a lower "sense of belonging to the field" and are less confident in their abilities in mathematics and economics. Furthermore, female students perceive a very high cost in tackling math tasks compared to male students (fear of getting low marks in math, nervousness and helplessness or anxiety). Concerning gender stereotypes, female students show a level of implicit stereotyping that is more than double that of male students (while their level of explicit stereotypes is very low). In particular, females with a high level of implicit gender stereotypes feel less confident in themselves and are not attracted by economics and mathematics. At the same time, they are less likely to enrol in economic degrees (preferring humanities careers). For males, the effect is the opposite: the higher the degree of stereotype, the higher their level of confidence in economics and mathematics.

Within our experiment, the treated group has been involved in a set of activities aimed at contrasting gender stereotypes (role models, board game, short films presentation), while the control group attended classes as usual without any change in their timetable. A descriptive analysis shows that female students - compared to the control group of female students - increase their awareness, and their propensity for STEM bachelor courses and show an increased preference for finance over marketing within the class of economics courses.

Our project also had a non-trivial impact on males. Treated male students appear to be more interested in pursuing higher education instead of looking for a job, are more uncertain about their future field of studies and increase their interest in humanities courses. In the field of economics courses, treated males show a higher propensity for marketing instead of finance when compared to the control group. Female students' explicit gender stereotypes increase after the experiment (from a very low level at the beginning). This could mean that our experiment has made female students more aware or freer to express themselves, or both. The same cannot be said for females' implicit gender stereotypes, which undergo only marginal variations. This result is not entirely unexpected, as implicit stereotypes are an inner construct that has developed throughout students' life and the time span of our experiment is too low to expect to obtain large impacts in this dimension. Satisfaction statistics reveal that for students - regardless of gender- the meeting with the female role models is the most useful and engaging among all proposed activities.

Although a careful investigation of the impact of the chosen activities on gender stereotypes and on college major choice would have required observing students along a longer time span than that of our experiment, our preliminary analysis has demonstrated that changes have occurred in the degree of (explicit and implicit) gender stereotypes, in the choice about enrolling in a university and in the choice about the field of study.

Further research on collected students' microdata includes the estimation of multivariate models to test the impact of the different determinants on college major choices, on students' confidence in their abilities and on college major choice itself.

Other extensions of our analysis leading to collect new microdata would be to measure teachers' and parents' gender stereotypes to test their impact and replicate the experiment in different areas of the country to test the impact of cultural factors and household models.

Since our microdata show that gender differences in parents' characteristics - are reflected in the gender distribution by type of high school attended and in the observable gender stereotypes that in turn affect college major choices, a relevant policy implication of our study would be to address gender inequalities in family background variables, within this perspective, the inclusion of gender equality as one of the three cross-cutting priorities in the Italian Recovery and Resilience Plan (Italian Government, 2021) can, if properly implemented, reduce the observed inequalities, though its impact could be observable only in the medium run. More specifically, policies addressing gender inequalities in high school and college major choices such as Summer Camp dedicated to girls to improve their skills in ICT or awareness programmes to reduce the impact of stereotypes on career and education choices should be promoted at the national, regional and local level (European Commission, 2021). In this regard, in the country analysed in this study, the Recovery and Resilience Plan includes a specific measure to improve STEM skills among female high school students (Italian Government, 2021, p.39), while at regional and local levels different measures to encourage choices of college majors free from gender stereotypes are detected (amongst them ‘Digital girls’, a summer camp for female students attending last years of high school to improve their ICT skills) (European Commission, 2021).

7. REFERENCES

AlmaLaurea. (2022). 'Laureate e Laureati: Scelte, Esperienze e Realizzazioni Professionali'. Bologna. https://www.almalaurea.it/sites/almalaurea.it/files/convegni/gennaio2022/6_almalaurea_rapportocompleto_laureatelaureati.pdf.

Astorne-Figari, C., and Speer Jamin D. (2019). 'Are Changes of Major Major Changes? The Roles of Grades, Gender, and Preferences in College Major Switching'. *Economics of Education Review* 70 (June): 75–93. https://doi.org/10.1016/j.econedurev.2019.03.005.

Avilova, T., Claudia Goldin, C. (2018). 'What Can UWE Do for Economics?' *AEA Papers and Proceedings* 108 (May): 186–90. https://doi.org/10.1257/pandp.20181103.

Bandura A. (1997). *Self-Efficacy: The Exercise of Control*. New York: Freeman.

Bandura, A., Barbaranelli C., Caprara G.V., and Pastorelli C. (2001). 'Self-Efficacy Beliefs as Shapers of Children's Aspirations and Career Trajectories'. *Child Development* 72 (1): 187–206. https://doi.org/10.1111/1467-8624.00273.

Breda, T., Grenet J., Monnet M., Van Effenterre C. (2020). 'Do Female Role Models Reduce the Gender Gap in Science? Evidence from French High Schools'. 13163. *IZA Discussion Papers*. IZA Discussion Papers. Institute of Labor Economics (IZA). https://ideas.repec.org/p/iza/izadps/dp13163.html.

Bertocchi, G., Luca B., and Marina M. (2022). 'Adams and Eves: The Gender Gap in Economics Majors'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4114415.

Booth, Alison L., Cardona-Sosa L., and Nolen P. (2018). 'Do Single-Sex Classes Affect Academic Achievement? An Experiment in a Coeducational University'. *Journal of Public Economics* 168 (December): 109–26. https://doi.org/10.1016/j.jpubeco.2018.08.016.

Bottazzi, L., and Lusardi L. (2020). 'Stereotypes in Financial Literacy: Evidence from PISA'. w28065. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w28065.

Carlana, M. , and Corno L. (2021). ‘Parents and Peers: Gender Stereotypes in the Field of Study’. 16582. *CEPR Discussion Papers*. CEPR Discussion Papers. C.E.P.R. Discussion Papers. https://ideas.repec.org/p/cpr/ceprdp/16582.html.

Carlana, M. (2019). ‘Implicit Stereotypes: Evidence from Teachers’ Gender Bias’. *The Quarterly Journal of Economics* 134 (3): 1163–1224. https://doi.org/10.1093/qje/qjz008.

Carrell Scott E., Marianne E. Page, and James E. West. (2010). ‘Sex and Science: How Professor Gender Perpetuates the Gender Gap’. *Quarterly Journal of Economics* 125 (3): 1101–44. https://doi.org/10.1162/qjec.2010.125.3.1101.

Carpenter, T. P., Pogacar R., Pullig C., Kouril M., Aguilar S., LaBouff J., Isenberg N., Chakroff, A. (2019). ‘Survey-Software Implicit Association Tests: A Methodological and Empirical Analysis’. *Behavior Research Methods* 51 (5): 2194–2208. https://doi.org/10.3758/s13428-019-01293-3.

Capalbi, A., (2015). La rappresentazione del lavoro femminile nell’archivio audiovisivo di Short on Work, in *Rappresentare il lavoro: percorsi interdisciplinari e crossmediali*, Quaderni della Fondazione Marco Biagi, QFMB Saggi, 5, 2015. https://www.fmb.unimore.it/rappresentare-il-lavoro-percorsi-interdisciplinari-e-crossmediali/

Capalbi, A., Giulia P. (2020). *Faccio Dunque Sono? Trasformazioni Del Lavoro e Dell’identità Di Genere Nelle Rappresentazioni Audiovisive Di Short on Work*. *SOCIOLOGIA DEL LAVORO*, no. 156 (March): 197–213. https://doi.org/10.3280/SL2020-156009.

Cimpian, J. R., Taek H. K., and Zachary T. McDermott. (2020). ‘Understanding Persistent Gender Gaps in STEM’. *Science* 368 (6497): 1317–19. https://doi.org/10.1126/science.aba7377.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 0 ed. New York: Routledge. https://doi.org/10.4324/9780203771587.

Corno, L. , La Ferrara E., Burns, J.. (2019). ‘Interaction, Stereotypes and Performance. Evidence from South Africa’. W19/03. *IFS Working Papers*. IFS Working Papers. Institute for Fiscal Studies. https://ideas.repec.org/p/ifs/ifsewp/19-03.html.

De Gioannis, E. (2022a). 'Implicit Gender-Science Stereotypes and College-Major Intentions of Italian Adolescents'. *Social Psychology of Education*, July. https://doi.org/10.1007/s11218-022-09709-3.

De Gioannis, E. (2022b). 'The Conundrum of Gender-Science Stereotypes: A Review and Discussion of Measurements'. *Quality & Quantity*, August. https://doi.org/10.1007/s11135-022-01512-8.

Doepke, M., Sorrenti, G., Zilibotti, F. (2019). The Economics of Parenting, *Annual Review of Economics*, 11 (1): 55–84

Duflo, E. (2004) 'Scaling Up and Evaluation | Innovations for Poverty Action'. https://www.poverty-action.org/publication/scaling-and-evaluation.

EIGE (2021) Gender Equality Index: Health. Vilnius, EIGE. Available at: https://eige.europa.eu/publications/gender-equality-index-2021-health

Eccles, J., and Wigfield A. (2002). 'Motivational Beliefs, Values, and Goals'. *Annual Review of Psychology* 53 (1): 109–32. https://doi.org/10.1146/annurev.psych.53.100901.135153.

Eccles, J. (1983). 'Expectancies, Values, and Academic Behaviors'. In *J. T. Spence (Ed.*), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: J. H. Freeman

European Commission (2021) *She Figures 2021, Gender in Research and Innovation Statistics and Indicators*, Luxembourg: Publications Office of the European Union.

Eurostat (2022) *Employment and activity by sex and age - annual data, LFSI_EMP_A*, Available at: https://ec.europa.eu/eurostat/databrowser/view/lfsi_emp_a/default/table?lang=en

Feather NT. (1992). 'Values, valences, expectations, and actions.' *J. Soc. Issues* (48):109–24

Feather NT. (1988). 'Values, valences, and course enrollment: testing the role of personal values within an expectancy-value framework.' *J. Educ. Psychol.* (80):381–91

Feld, J., and Zolitz, U. (2022). 'The Effect of Higher-Achieving Peers on Major Choices and Labor Market Outcomes'. *Journal of Economic Behavior & Organization* 196 (April): 200–219. https://doi.org/10.1016/j.jebo.2022.01.012.

Frome, P. M., & Eccles, J. S. (1998). 'Parents' influence on children's achievement-related perceptions'. *Journal of Personality and Social Psychology, 74*(2), 435–452. https://doi.org/10.1037/0022-3514.74.2.435

Giustinelli, P. (2016). 'Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of the High School Track', *International Economic Review*, 57 (2): 573–602, https://doi.org/10.1111/iere.12168.

Greenwald, A. G., Nosek B. A., and Banaji M. R. (2003). 'Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm.' *Journal of Personality and Social Psychology* 85 (2): 197–216. https://doi.org/10.1037/0022-3514.85.2.197.

Greenwald, A. G., and Nosek, B. A. (2001). 'Health of the Implicit Association Test at Age 3.' *Zeitschrift Für Experimentelle Psychologie* 48 (2): 85–93. https://doi.org/10.1026/0949-3946.48.2.85.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). 'Measuring individual differences in implicit cognition: The implicit association test'. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464.

Good, C., Rattan, A., & Dweck, C. S. (2012). 'Why do women opt out? Sense of belonging and women's representation in mathematics'. *Journal of Personality and Social Psychology, 102*(4), 700–717. https://doi.org/10.1037/a0026659 .

Goldin C. (2015). Gender and the undergraduate Economics major: Notes on the Undergraduate Economics major at a highly selective liberal arts college. Mimeo, Harvard University.

Ingellis A. G., Riccardi F., Cristini C., and Riccardi F., (2018). 'Mind the Gap - Gender Gaps in the Education-to-Work Transition in Mediterranean Countries. The Role of Gender Stereotypes in Youngsters' Perception'. ResearchGate. https://www.researchgate.net/publication/340918236_Mind_the_Gap.

Italian Government (2021). *Piano Nazionale di Ripresa e Resilienza (National Recovery and Resilience Plan)*, (https://www.governo.it/sites/governo.it/files/PNRR.pdf

Jacob, B., and Lars Lefgren L. (2005). Principals as Agents: Subjective Performance Measurement in Education. w11463. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w11463.

Jethwani, M. M., Memon, N. Seo W., and Richer, A. (2017). '"I Can Actually Be a Super Sleuth": Promising Practices for Engaging Adolescent Girls in Cybersecurity Education'. *Journal of Educational Computing Research* 55 (1): 3–25. https://doi.org/10.1177/0735633116651971.

Lane, Kristin A., Jin X. Goh, and Erin Driver-Linn. 2012. 'Implicit Science Stereotypes Mediate the Relationship between Gender and Academic Participation'. *Sex Roles* 66 (3–4): 220–34. https://doi.org/10.1007/s11199-011-0036-z.

Lavy, V. (2008). 'Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment'. *Journal of Public Economics*, 92 (10-11): 2083-2105. doi:10.1016/j.jpubeco.2008.02.009

Lizzeri, A., Siniscalchi, M. (2008). 'Parental Guidance and Supervised Learning', *The Quarterly Journal of Economics*, 123 (3): 1161–1195. https://doi.org/10.1162/qjec.2008.123.3.1161

Martin, C. L., and Dinella, L. (2001). 'Gender-related development'. *International Encyclopedia of the Social & Behavioral Sciences* (pp. 6020–6027). Elsevier.

Megalokonomou R., Vidal-Fernández M., and Yengin D. (2021). 'Underrepresentation of Women in Undergraduate Economics Degrees in Europe: A Comparison with STEM and Business'. *IZA - Institute of Labor Economics* Policy Paper No. 175. https://www.iza.org/de/publications/pp/175/underrepresentation-of-women-in-undergraduate-economics-degrees-in-europe-a-comparison-with-stem-and-business.

Merritt, S. K., Hitti A., Van Camp A. R., Shaffer E., Sanchez M. H., and O'Brien L. T. (2021). 'Maximizing the Impact of Exposure to Scientific Role Models: Testing an Intervention to Increase Science Identity among Adolescent Girls'. *Journal of Applied Social Psychology* 51 (7): 667–82. https://doi.org/10.1111/jasp.12774.

Nosek, Brian A., Anthony G. Greenwald, and Mahzarin R. Banaji. (2005). 'Understanding and Using the Implicit Association Test: II. Method Variables and Construct Validity'. *Personality and Social Psychology Bulletin* 31 (2): 166–80. https://doi.org/10.1177/0146167204271418.

OECD. (2021). *Education at a Glance 2021: OECD Indicators*. OECD. https://doi.org/10.1787/b35a14e5-en.

OECD. (2019). *PISA 2018 Results* (Volume II): *Where All Students Can Succeed*, *PISA*, OECD Publishing, Paris, https://doi.org/10.1787/b5fd1b8f-en .

Paredes Fuentes S., Arzhevikina E., Bentham C., Buratta. (2021). 'How can we promote diversity in Economics?' *The Economics Observatory*. https://www.economicsobservatory.com/how-can-we-promote-diversity-in-economics .

Patnaik, Arpita, Matthew Wiswall, and Basit Zafar. (2020). College Majors. w27645. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w27645.

Porter, C., and Serra, D. (2020). 'Gender Differences in the Choice of Major: The Importance of Female Role Models'. *American Economic Journal: Applied Economics* 12 (3): 226–54. https://doi.org/10.1257/app.20180426.

Redmon, P. and Gutke, H. (2020). 'STEMming the Flow: Supporting Females in STEM'. *International Journal of Science and Mathematics Education* 18 (2): 221–37. https://doi.org/10.1007/s10763-019-09963-6.

Rokeach, M. (1979). 'From the individual to institutional values with special reference to the values of science.' In M. Rokeach (Ed.), Understanding human values (pp. 47-70). New York, NY: Free Press.

Shan X. (2020). Does minority status drive women out of male-dominated fields. mimeo.

Stoeger, H., Schirner S., Laemmle L., Obergriesser S., Heilemann M., and Ziegler, A. (2016). 'A Contextual Perspective on Talented Female Participants and Their Development in Extracurricular STEM Programs: Talented Females in an Extracurricular STEM Program'. *Annals of the New York Academy of Sciences* 1377 (1): 53–66. https://doi.org/10.1111/nyas.13116.

Tiedemann, J. (2000). ‘Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school’. *Journal of Educational Psychology, 92*(1), 144–151. https://doi.org/10.1037/0022-0663.92.1.144

Tomasetto C., Mirisola A., Galdi S., and Cadinu, M. (2015). ‘Parents’ Math–Gender Stereotypes, Children’s Self-Perception of Ability, and Children’s Appraisal of Parents’ Evaluations in 6-Year-Olds’. *Contemporary Educational Psychology* 42: 186–98. https://doi.org/10.1016/j.cedpsych.2015.06.007.

APPENDIX 1

IAT for gender and field of study

a. How IAT works

The test measures mental association by looking at the speed of stimuli classification.

Stimuli belong to two pairs of categories which are formed on opposite hemispheres: table 1 reported the categories and stimuli that we chose (Female vs Male) and (Eco-fin vs Humanistic).

Classification is done by placing stimuli to the right[12] or left[13] depending on the category they belong to and the exercise request.

The setting and implementation (next paragraph) of the IAT that we adopt are inspired by Carpenter et. al. 2019, who, in addition to building a free IAT software, tested its validity with three different experiments. The test consists of 7 blocks in which there are stimuli belonging to the four different categories; the task is to classify them as faster as possible while avoiding mistakes.

Blocks 1, 2 and 5 are practice blocks in which appear only names (target A and B) or subjects (attribute A and B), meanwhile blocks 3,4,6,7 are combined in the sense that appears both people's names and subjects. Half of the 4 blocks require working on *"compatible tasks"* by classifying male & economics and finance on one side and female and humanistic on the other. Meanwhile, in the remaining blocks, the pairs are reversed: male & humanistic stored on one side and female and economics and finance on the other "*incompatible tasks"***.**

This setting is repeated four times in which categories have different orders and places (table 2), and **only one** of the four permutations is randomly assigned to each respondent. According to Nosek et al., 2005 this randomised permutation – aimed at counterbalancing the left/right starting positions of targets and attributes - allows for more precise estimates as is proved that the order of task (blocks) matters in IAT performance (Greenwald & Nosek, 2001).

The data originating from the reply in blocks 3,4,6,7 are used to compute standardised differences scores (Dscore)[14] in IAT, indicating in which condition (compatible vs. incompatible) participants were faster. Is reliable to assume (and was proved) that a person suffering from stereotype is quicker to perform the compatible task and has more difficulty in the incompatible one: since he mentally associates females with the humanistic hemisphere and boys with the economic one. The Dscore is positive (Dscore > 0) for people who are faster at compatible tasks revealing the presence of stereotypes while a negative one (Dscore < 0) discloses the opposite mental association (female associated with economics and finance and male with humanities).

---

[12] By pressing the "I" key on the keyboard.
[13] By pressing the "E" key on the keyboard.
[14] The Dscore calculation follows Greenwald et al. (2003).

The index we get is a relative measure: that means that we cannot disclose if females are associated with the humanistic field and males with the economic one in an absolute sense. Instead, we can keep which gender is seen relatively closer to the economics/humanistic hemisphere at the implicit level.

Table A1- IAT stimuli

Categories	Stimuli
Masculine names (target A)	Luca, Federico, Matteo, Alberto, Davide, Alessandro.
Feminine names (target B)	Anna, Martina, Laura, Giulia, Erica, Alessia
Economics and finance (attribute A)	Economics, finance, calculus, statistics, mathematics, algebra, formulas, equations).
Humanistic (attribute B)	literature, Italian, history, philosophy, art, pedagogy, languages

Table A2- IAT blocks

Block	Compatible First [Target A on Right with Pos]		Incompatible First [Target A on Right with Neg]		Compatible First [Target A on Left with Pos]		Incompatible First [Target A on Left with Neg]	
	Right	**Left**	**Right**	**Left**	**Right**	**Left**	**Right**	**Left**
1	F	M	F	M	M	F	M	F
2	H	E	E	H	E	H	H	E
3	F+H	M+E	F+E	M+H	M+E	F+H	M+H	F+E
4	F+H	M+E	F+E	M+H	M+E	F+H	M+H	F+E
5	M	F	M	F	F	M	F	M
6	M+H	F+E	M+E	F+H	F+E	M+H	F+H	M+E
7	M+H	F+E	M+E	F+H	F+E	M+H	F+H	M+E

Notes:
M: stands for Masculine names (target A) *F: stands for Feminine names (target B)*
E: stands for Economics and Finance (attribute A) *H: stands for Humanistic (attribute B)*

b. How we implemented it

The typology and fineness of the information to be collected (such as milliseconds reaction times) implies the need for a computer to administer the questionnaire and a well-reasoned procedure to implement it. The latter is because common online survey tools are unable to detect sensitive measures of reaction time.

For this, we use the software developed by Carpenter et. al. 2019: authors created an open-source tool (IATgen[15]) which could be easily customized to research needs and returns Qualtrics-compatible output: this is a great advantage as it allows the IAT test to be conducted directly within a longer questionnaire and thus permits linking it to any other desired respondent information.
Out procedure is detailed in the following steps.
At first, we select our stimuli after delving into literature (Carlana 2019), consulting with professor Carlo Tomasetto and conducting a pilot survey[16] on students of the Department of Economics Marco Biagi of the University of Modena and Reggio Emilia and of the Department of Psychology Renzo Canestrari of the University of Bologna.
Then, we insert the chosen stimuli in the IATgen interface grouped into the four categories (two targets and two attributes) and download the IAT test in the QSF format, which could be imported directly into paid Qualtrics account.
But before importing it, through HTML codes, we carefully translated into Italian the instruction of the test that appears to users. This is a crucial point of our implementation because it is <u>not</u> suggested to make any changes to the test once in Qualtrics because the data collection will get corrupted[17].
The questionnaire begins with a barrier that verifies the device used and excludes those who do not have a keyboard because they are unable to conduct the IAT, follows with randomization that assigns an IAT permutation to each respondent, and ends with all other questions related to our covariates of interest.
Finally, we test the questionnaire to be sure that everything works, and that data are stored correctly and could be imported into IATgen.

c. Data cleaning and imputation of missing values

We have 10 missing in the score in the Implicit Association Test: missingness is imposed by the software and arises when it detects that the test was done without diligence and effort. Namely following the procedure firstly reported by Greenwald et. al., 2003 and resumed in Lane et al., 2007 (p. 92). Are scored as missing trials too slow (over 10,000 ms) and trials in which participants place more than 10% of the stimuli faster than 300 ms.
So, we cannot rely upon the missing values in the IATest to be completely random (MCAR) but, instead, related to the characteristics of the respondents (MAR) as the low motivation of respondents to carry out the exercise or a deficit in understanding it.

---

[15] http://iatgen.org/
[16] In the pilot survey we subjected students to a prototype of the IAT, but with a larger pool of options (namely Economics, Finance, Calculus, Statistics, Mathematics, Algebra, Formulas and Equations as economics and finance stimuli & Humanities, Italian, Literature, History, Philosophy, Art, Pedagogy, Languages as humanistic stimuli). Following we explicitly ask them to evaluate all these options from 1 (= highly humanistic domain) to 10 (=highly economics domain) To check that the stimuli chosen are truly perceived by pupils as belonging to the two hemispheres (economics and humanities) and select those that are sensed as most representative.
[17] This is particularly insidious since the error is initially invisible (the questionnaire works correctly as the data collection and storage), damages arise only once the data are reimported into IATgen.

APPENDIX 2

Description of the main variables in the data set

Table A3- Description of the main variables

VARIABLE	DEFINITION
Dependent variables	
Finance Marketing STEM Humanistic Work	Continuous variables, ranging from 1 “Extremely Unlikely” to 5 “Extremely Likely”, which show how likely students are to take the following pathways: *Finance:* enrol in economics and finance degree. *Marketing:* enrol in economics and marketing degree. *STEM:* enrol in a “*hard STEM degree*”, namely math or statistics or physics or chemistry or engineering or architecture and urban planning. *Humanistic:* enrol in a humanistic degree, namely political science, sociology, law, philosophy, languages, pedagogy and psychology. *Work:* go to work. For composite groups, the variable is the average between all options.
Student observable characteristics	
Female	Binary variable taking value 1 if female, 0 if male.
Age	continuous variable calculated as the difference between the year of the testing (2022) and the student’s birth.
YearsInEmiliaRomagna	continuous variable indicating the years passed in the Emilia Romagna region. This variable coincides with the age of the students who were born in the region.
North	Binary variables taking value 1 for students born in North[18] of Italy and 0 for students from the rest of Italy or abroad.
RepetedClass	Binary variables taking value 1 for students have ever repeated a grade or more
Environmental Factors	
Family Background	

[18] We define North as the ITC and ITH NUTS1 (Nomenclature of Territorial Units for Statistics) codes of Italy. Namely Piemonte, Valle D’Aosta, Liguria and Lombardia for ITC NUTS1 (North West Italy) and Trentino Alto Adige, Veneto, Friuli Venezia Giulia, Emilia Romagna for ITH NUTS1 (North East Italy).

Mother_North	Binary variables taking value 1 if the student's mother was born in the North of Italy and 0 for mothers from the rest of Italy or abroad.
Mother' Level Of Study: Mother_Elementarymiddle Mother_Highschool Mother_Universitymore	Binary variables represent the mother's education level. *ElementaryMiddle* takes value 1 if the mother completed no more than lower secondary school, *highSchool* takes value 1 if she went to high school, *universityMore* takes value 1 whether he attended a bachelor's, master's or doctoral degree course.
Mother Education. High School Spec: Mother_Humanisticlyceum Mother_Scientificlyceum Mother_Technical Mother_Vocational	Set of binary variables specifying the nature of the *high school* attended by the mother. The four dummies respectively have value 1 whether the mum attended a humanistic Lyceum (*Mother_Humanisticlyceum*) , a scientific lyceum (*Mother_Scientificlyceum*), a technical school (*Mother_Technical*) or a vocational school (*Mother_Vocationa*l) and 0 otherwise (others levels of study).
Mother Degree: Mother_ECO Mother_STEM Mother_HUM	Set of binary variables specifying the nature of the *degree course* attended by the mother. The three dummies respectively have value 1 whether the mum attended a humanistic degree (*MOTHER_HUM*), a economics degree (*MOTHER_ECO*), a stem degree (*MOTHER_STEM*) and 0 otherwise (others levels of study or "neutral" degree field do not strictly enter in humanities, economics or stem).
Mother Work: Mother_Housewife Mother_Notemployed Mother_Empl1 Mother_Empl2 Mother_Empl3 Mother_Empl4	Binary variables representing the mother's current employment status. *MOTHER_housewife* takes value 1 if the mother is a housemaker. *MOTHER_notEmplyed* takes value 1 if the mother doesn't work. The other dichotomies indicate a labour status which increases with the number, namely: *MOTHER_empl1* takes value 1 for: blue collar, atypical worker (as a term-contract worker, an occasional collaborator etc...) and low-paying jobs (waiter, school janitor, gardener, social worker, cleaner, bartender, caregiver …) *MOTHER_empl2* takes value 1 for employee, teacher, nurse, physiotherapist or for the self-employed worker as a trader, shopkeeper or artisan. *MOTHER_empl3* takes value 1 for freelance professionals (as a lawyer, commercialist, etc..), university professor, headmaster. *MOTHER_empl4* takes value 1 for middle manager, doctor, surgeon, partner and or owner of a company.
Father_North	Binary variables taking value 1 if the student's Father was born in the

	North of Italy and 0 for fathers from the rest of Italy or abroad.
Father' Level Of Study: Father_Elementarymiddle Father_Highschool Father_Universitymore	Binary variables represent the father's education level. *ElementaryMiddle* takes value 1 if the father completed no more than lower secondary school, *highSchool* takes value 1 if she went to high school, *universityMore* takes value 1 whether he attended a bachelor's, master's or doctoral degree course.
Father Ed. High School Spec: Father_Humanisticlyceum Father_Scientificlyceums Father_Technical Father_Vocational	Set of binary variables specifying the nature of the *high school* attended by the father. The four dummies respectively have value 1 whether the Dad attended a humanistic Lyceum (*Father_Humanisticlyceum*) , a scientific lyceum (*Father_Scientificlyceums*), a technical school (*FATHER_Technical*) or a vocational school (*FATHER_Vocational*) and 0 otherwise (others levels of study).
Father Degree: Father_ECO Father_STEM Father_HUM	Set of binary variables specifying the nature of the *degree course* attended by the father. The three dummies respectively have value 1 whether the Dad attended a humanistic degree (*FATHER_HUM*), a economics degree (*FATHER_ECO*), a stem degree (*FATHER_STEM*) and 0 otherwise (others levels of study or "neutral" degree field do not strictly enter in humanities, economics or stem).
Father Work: Father_Notemployed Father_Empl1 Father_Empl2 Father_Empl3 Father_Empl4	Binary variables representing the father's current employment status. *FATHER_notEmplyed* takes value 1 if the father doesn't work (there are no fathers housemaker in the sample). The other dichotomies indicate a labour status which increases with the number, namely: *FATHER_empl1* takes value 1 for: blue collar, atypical worker (as a term-contract worker, an occasional collaborator etc...) and low-paying jobs (waiter, school janitor, gardener, social worker, cleaner, bartender, caregiver …) *FATHER_empl2* takes value 1 for employee, teacher, nurse, physiotherapist or for self-employed worker as a trader, shopkeeper or artisan. *FATHER_empl3* takes value 1 for freelance professionals (as a lawyer, commercialist, etc..), university professor, headmaster. *FATHER_empl4* takes value 1 for the middle manager, doctor, surgeon, partner and or owner of a company.
MumCare	Continuous variable ranging from 1 "never" to 5 "several times a month" reporting the time and attention that the mother devotes to the student (including speaking about his future, the importance of school

	and asking how things are going in the institute. Composite index inspired by PISA 2015.
DadCare	Continuous variable ranging from 1 "never" to 5 "several times a month" report the time and attention that the father devotes to the student (including speaking about his future, the importance of school and asking how things are going in the institute. Composite index inspired by PISA 2015.
WellnessStudy	Continuous variable ranging from 0 to 5 indicating the possession of some “study assets” like a PC, dictionary, internet connection, room and a table intended for study. Composite index inspired by PISA 2012.
WellnessHome	Continuous ranging from 0 to 6 indicating the possession of some basic “home assets” as dishwasher, fridge, oven, etc... Composite index inspired by PISA 2012.
I.Bookathomenomissing	Categorical variable indicating the estimated number of books at home. Inspired by PISA 2012.
School Environment	
Type Of School: Lyceum Technical Vocational	Binary variables corresponding to the three main Italian tracks of upper secondary schools: Lyceums, Technical, and Vocational.
ClassSize	Continuous variable representing the number of pupils in the classroom.
ShareFemale	Continuous variable equal to the percentage of females in the class[19].
FemaleTeacher	Binary variable indicating whether students have a female maths or economics teacher.
TeacherBeliefInTheProject	Continuous variable which is reported the mean of the staff’s evaluation about how much they feel professors believe in the usefulness of the project. In your opinion, did ***school X***'s professor(s) believe in the usefulness of the project?
TeacherMotivateStudent	Continuous variable which is reported the mean of the staff’s evaluation about how much they feel professors make their classes

[19] The percentage was computed on respondents to the pre questionnaire (93% of the total number of participants), so could slightly differ from the real one.

	get involved in the project. In your opinion, how motivated was(were) ***school X***'s professor(s) to make his(their)class(es) get involved in the project?
TeacherEngagement	Continuous variable reporting the mean of the staff's evaluation about how much they feel professors satisfy to participate in the project. how satisfied/happy you felt ***school X***'s professor(s) was(where) that his(their)class(es) was participating in the project?
OnlineRoleModel	Binary variable equal to 1 if the role model activity was conducted online.
OnlineShort	Binary variable equal to 1 if the short film projection activity was conducted online.
School	Binary variables corresponding to the six schools participating in the projects.
Class	Binary variables corresponding to the classes participating in the projects.
TimeSpan	Continuous variable indicating the number of days between the first and second questionnaires.
Hmath	Sum of weekly math hours in the curriculum (first through fifth grades)
Heco	Sum of weekly economy hours in the curriculum (first through fifth grades)
Choices Determinants	
Stereotype	
Dscore	Continuous variable indicating the Implicit stereotype (composite index based on the Iat test results). The variables range from -1.04 to + 1.26. Positive values indicate a positive mental association between female&humanistic and male&finance/economy, neutral values reflect the absence of stereotypes and negative values the opposite mental association (male&humanistic and female&finance/economy).
ExplicitstereotypeSchool	Continuous variables ranging from 1 "no stereotype" to 6 "high stereotype" indicating the female-economy explicit association at school. (composite index)
ExplicitstereotypeFamily	Continuous variables ranging from 1 "no stereotype" to 6 "high stereotype" indicating the female-economy explicit association in the household. (composite index)
Expectation Of Success	
MeasuredAbility	Continuous variable ranging from 4 to 10 correspond to the average

	grades in math and economics took by students at the end of the fourth grade.
AbilitySelfConcept	Continuous variable ranging from 1 “low self-confidence in math” to 6 high “high self-confidence in math”. The variable is composed of dimensions like the student agreement with the following statements “I am not good at math (reversed)”, “I get good grades in mathematics”, “I learn mathematics quickly”, “In my maths class, I understand even the most difficult work”. Composite index built on some items extrapolated from a battery of PISA questions. (Alpha subindex pre-aggregation = Scale reliability coefficient: 0.84)
Overconfidence_Math	Binary variable equal to one whether teens have a perception that they are better at math than what their actual grade reveals and zero otherwise.
Overconfidence_Economy	Binary variable equal to one whether teens have a perception that they are better at economy than what their actual grade reveals and zero otherwise.
Subjective Tasks Value	
Like_Maths	Continuous variable ranging from 1 to 6 indicating how much students like the math subjects.
Like_Economics	Continuous variable ranging from 1 to 6 indicating how much students like the economy subjects. Even if this variable is highly predictive and correlated with the choice of the field of study we cannot use it in regression because
ImpoeconomyForLife	Continuous variable ranging from 1 “not important” to 5 “extremely important” which corresponds to how important students perceive the economy to be for their life.
ImpoeconomyForAccademy	Continuous variable ranging from 1 “not important” to 5 “extremely important” which corresponds to how important students perceive the economy to be for their academic future.
ImpoeconomyForProfession	Continuous variable ranging from 1 “not important” to 5 “extremely important” which corresponds to how important students perceive the economy to be for their professional future.
Cost	Continuous variable ranging from 1 “low perceived cost” to 6 “high perceived cost” (we mean as costs feelings such as fear to get low grades in math, nervousness and|or helpless and|or anxiety in solving math problems and exercises) . Composite index built on some items extrapolated from a battery of PISA questions.

Senseofbelongingindex1 Senseofbelongingindex2	Continuous variable ranging from 1 “low sense of belonging to financial-economic community” to 8 “high sense of belonging to financial-economic community” The variables included in the index explore the feeling of the student (who is asked to imagine him/herself in an economics and finance class) S[…]x1 - subscale acceptance: I feel like an outsider/accepted/ respected/disregarded/valued/ neglected/appreciated. S[…]x2 - subscale membership: I feel a member of the economy and finance world/ that I belong to the economy and finance community/ in connection with the economy and finance community. Composite index inspired on the reduced scale (and adapted to the financial-economic domain) of Good et all 2012.

Table A4 – Treatment, details on heterogeneity

INSTITUTE	CLASS(ES)	FREE TO CHOOSE	SHORT ON WORK	ROLE MODEL
Vocational1	5CP	**Date**: 25/11/2021 **Mode:** presence **Staff:** Staff1[1]. ; Staff2[2]	**Date:** 27/11/2021 **Mode:** presence **Staff:** Staff1 ; Staff2	**Date**: 02/12/2021 **Mode:** presence **Staff:** Staff1 ; Staff2 **Role model:** Role model1[3]
Vocational2a	5A	**Date**: 25/11/2021 **Mode:** presence **Staff:** Staff1 ; Staff2	**Date:** 02/12/2021 **Mode:** presence **Staff:** Staff1 ; Staff2	**Date**: 09/12/2022 **Mode:** Role model online, students and staff in class. **Staff:** Staff1; Staff2; Staff3 [2] **Role model:** Role model2 [4]
Lyceum1	5A ES 5B ES	**Date**: 29/11/2021 **Mode:** presence **Staff:** Staff1 ; Staff2	**Date:** 01/12/2021 **Mode:** Staff online, students in class. **Staff:** Staff1 ; Staff2	**Date**: 04/12/2021 **Mode:** presence **Staff:** Staff3 **Role model:** Role model3 [5]
Technical1a	5A AFM 5B AFM	**Date**: 17/01/2022 **Mode:** presence **Staff:** Staff1; Staff2	**Date:** 20/01/2022 **Mode:** presence **Staff:** Staff1; Staff4 [2]	**Date**: 22/01/2022 **Mode:** presence **Staff:** Staff3 **Role model:** Role model4 [6]
Technical1a	5F SIA	**Date**: 17/01/2022 **Mode:** presence **Staff:** Staff1; Staff2;	**Date:** 20/01/2022 **Mode:** presence **Staff:** Staff3 ; Staff4	**Date**: 22/01/2022 **Mode:** presence **Staff:** Staff3 **Role model:** Role model4
Technical1b	5H RIM 5L RIM	**Date**: 18/01/2022 **Mode:** presence **Staff:**	**Date:** 25/01/2022 **Mode:** presence **Staff:** Staff1 ; Staff5[2]	**Date**: 29/01/2022 **Mode:** Role model online, students and staff in class. **Staff:** Staff2 **Role model:** Role model5 [7]
Vocational2b	5M	**Date**: 24/01/2022 **Mode:** presence **Staff:** Staff2; Staff1	**Date:** 01/02/2022 **Mode:** presence **Staff:** Staff1 ; Staff2	**Date**: 04/02/2022 **Mode:** presence **Staff:** Role model online, students and Staff2 in class. **Role model:** Role model6 [8]

continued on the next page

INSTITUTE	CLASS(ES)	FREE TO CHOOSE	SHORT ON WORK	ROLE MODEL
Lyceum2	5B CLAS	**Date**: 27/01/2022 **Mode:** presence **Staff:** Staff1	**Date**: 03/02/2022 **Mode:** presence **Staff:** Staff1; Staff4	**Date**: 12/02/2022 **Mode:** presence **Staff:** Staff1 **Role model:** Role model7 [9]
Lyceum2 (due to the Pandemic issue -for this class- "Free to Choose" and "Short on Work" were inverted in the implementation timeline)	5A LING	**Date**: 03/02/2022 **Mode:** presence (due to Pandemic issue "Free to Choose" came after "Short on Work") **Staff:** Staff2	**Date**: 27/01/2022 **Mode:** online with students at home (due to Pandemic issue Short on Work") came before "Free to Choose") **Staff:** Staff2	**Date**: 12/02/2022 **Mode:** presence **Staff:** Staff1 **Role model:** Role model7
Technical2	5B RIM	**Date**: 10/01/2022 **Mode:** presence **Staff:** Staff3; Staff2	**Date:** 19/01/2022 **Mode:** online with students at home **Staff:** Staff1	**Date**: 28/01/2022 **Mode:** presence **Staff:** Staff1 **Role model:** Role model1

[1] Research fellow at DHMore (Interdepartmental Center on Digital Humanities, University of Modena and Reggio Emilia).

[2] PhD student in Labour, Development and Innovation at Department of Economics Marco Biagi and Economic Policy Foundation.

[3] PhD student in Labour, Development and Innovation at Department of Economics Marco Biagi and Economic Policy Foundation, ex-worker in the finance section of Prometeia. Master's degree in Economics and Finance at Marco Biagi Department of Economics.

[4] Research Assistant at the Department of Communication and Economics, University of Modena & Reggio Emilia. PhD's degree in Labour Relations at Department of Economics Marco Biagi and Economic Policy Foundation.

[5] Economist in the finance section of Prometeia. Master's degree in Economics and Public Policies at Marco Biagi Department of Economics.

[6] Data and analysis manager of the Job Pricing observatory. PhD's degree in Economics and business at the University of Cagliari.

[7] Associate Professor in Political Economy SECS/P01. PhD's degree in Economics at London School of Economics and Political Science.

[8] Senior economist in the finance section of Prometeia. Master's degree in Economic and Social Sciences at Bocconi university.

[9] Operating assistant in a financial advisory firm of Banca Fideuram. Master's degree in Labour Relations at Marco Biagi Department of Economics

ESSAY N° 2

The role of stereotypes, the subjective task value and the expectation of success in college major choices

Chiara Tasselli[a] , Tindara Addabbo[b]

a University of Modena and Reggio Emilia, Marco Biagi Foundation
Email: chiara.tasselli@unimore.it

b University of Modena and Reggio Emilia, Gender Equality Observatory, Marco Biagi Foundation
Address: Viale Berengario 51, 41121, Modena, Italy
Email: tindara.addabbo@unimore.it

The author gratefully acknowledges the valuable guidance provided by Prof. Jaya Krishnakumar of the (Institute of Economics and Econometrics - Université de Genève) for carrying out this research.

The role of stereotypes, the subjective task value and the expectation of success in college major choices.[20]

ABSTRACT

Gender disparities in college major choices perpetuate inequalities in the labour market, disadvantaging women. Previous research has highlighted the impact of gender stereotypes on college major choices and the individual and societal costs associated with a mismatch between talents and educational fields.

This study uses Eccles' Expectancy Value Model to identify key drivers of college major choices among high school students and examine how they are influenced by stereotypes.

Data were collected during an experimental project involving 26 classes from 6 high schools in Northern Italy. Stereotypes were measured through the Implicit Association Test (IAT) and sets of indicators were aggregated to derive five main choice determinants: namely extrinsic and intrinsic motivation, costs, sense of belonging to the field and the ability self-perception. According to Eccles's model, a generalized structural equation model (GSEM) is specified to disentangle the relationships between stereotypes, choice drivers, and actual choices. IAT results confirm - especially for females - a nonnegligible presence of implicit stereotyping. The GSEM application reveals that stereotypes negatively affect almost all subjective values associated with the mathematical and economic domains, with the exception of costs such as anxiety, nervousness, and the belief of not succeeding, which impact female choices regardless of stereotype levels. Additionally, intrinsic motivation and ability self-concept affect the gap between economic-mathematical and humanistic skills, while the sense of belonging, extrinsic and intrinsic motivation boost enrolment intentions towards universities of economics, math, and statistics. In conclusion, our findings suggest that it's not a biological matter of gender that steers women away from quantitative fields of study, but rather the stereotypes they are burdened with that influence indirectly their choices and performance passing through their subjectives beliefs. Therefore, policies and interventions that aim to combat stereotypes and promote a more inclusive and equitable learning environment could help address the gender gap in academic choices and performance. Overall, this study highlights the need to address gender stereotypes in college major choices to promote greater gender balance in college pathway choices to facilitate equality in the labor market.

[20] The author gratefully acknowledges the valuable guidance provided by Prof. Jaya Krishnakumar of the (Institute of Economics and Econometrics - Université de Genève) for carrying out this research.

0. INTRODUCTION

Gender differences in university choice underlie the widespread underrepresentation of women in Europe in math-intensive undergraduate and doctoral programs. This difference is even more pronounced in Mediterranean countries and familyist cultures such as Italy: the report of the XXIV Almalaurea's survey on the profile of Italian college graduates reveals a large imbalance across gender and field of study. In bachelor courses, women are a conspicuous majority in the education (93.1%), linguistics (85.1%), psychology (81.5%) groups and this percentage slightly raises in the master course[21]. On the other hand, they shy away from the quantitative field and they are underrepresented in computer science and technology (ICT < 15%), engineering (< 26%) (Almalaurea 2022) and high quantitative domain.

The unbalance gender distribution across fields of study results in a female disadvantage in the labour market. While higher earnings and better career opportunities appear to be related to male-dominated majors such as hard STEM and finance, female-dominated majors (education, humanities and a set of social sciences) are characterized by lower work opportunities and lower salaries (Carlana & Corno 2021, Carlana 2019, Cimpan 2020, Goldin 2015, Ingellis et all 2018). Thus, persistence in gender differences in fields of study is bound to be reflected in persistence in gender inequalities in the labour market (OECD, 2021). In fact, another report by Almalaurea (the report on the employment status of graduates) reveals that students from humanistic fields - once in the labour market - suffer from both higher unemployment and lower wages. Data for bachelor's students reveals that the highest percentages of unemployed are observed in art and design (26.9 %), linguistics (23.7 %), political-social and communication (19.9 %), literary-humanities (19.1 %) and psychology (18.9 %), moreover, the graduates from this fields are also those with the lowest wages earning on average less than 1100 euros per month[22] (Almalaurea 2022b). Needless to add that the same report shows unemployment rates among the lowest and wages among the highest for those degree programs with a low concentration of women, such as engineering, computer science and highly quantitative courses.

The concentration of women in humanities paths is an important issue, particularly given the existing disadvantages that women face in the labour market. Due to biological factors and patriarchal culture, women tend to have lower labour-force participation, experience, working hours, and training hours, leading to lower salaries and vertical segregation. In this study, we aim to shed light on the causes of unequal gender distribution in fields of study as the main cause of horizontal segregation.

---

[21] Also due to the increase of female share in master courses.

[22] One year after graduation.

A particular focus is on how subjective beliefs (choices determinants) are influenced by stereotypes and how this impacts choices.

Special attention is given to inner subjective beliefs: how they are affected by stereotypes and, in turn, their consequences on choices.

To analyse the determinants of college major choices, we collected primary data from 26 classes in six high schools in two Northern districts of Italy and analysed the reported preferences of students in their last year of studies, which is the period immediately preceding college major choice. Stereotypes were detected through the Implicit Association Test (IAT), which measures automatic beliefs about the association between gender and a given field of education. Inner subjective beliefs were defined through Eccles' theory, which posits that choices and performances are strongly guided by intrinsic and extrinsic motivation, the perceived cost, the sense of belonging, and the ability-self-perception.

While there is already broad theoretical and empirical support for the presence of implicit stereotypes, their causes, and their impact on choices, less explored is the mechanism through which gender stereotypes affect educational choices. Our research contributes to the literature on the determinants of college major choices in three ways: by focusing on Italy, by analyzing students' enrollment intentions rather than their actual choices, and by disentangling the inner subjective mechanism that leads to biased choices influenced by stereotypes.

The paper is structured as follows: Section 1 analyzes the relevant literature and outlines the research questions. Section 2 illustrates the theoretical framework derived from Eccles' theory, which serves as the basis for the GSEM models developed in Section 5. The results of these models are presented in Section 6. Section 3 provides a description of the data and variables used in the study, while Section 4 presents descriptive statistics. Section 7 summarizes the findings and provides policy implications. Finally, Sections 8 and 9 contain the bibliography and annexes, respectively.

1. LITERATURE REVIEW[23]

Differences in college major choices by gender are at the origin of the widespread underrepresentation of women in math-intensive majors and doctoral courses (OECD, 2019, 2021; European Commission, 2021). While higher earnings and better career opportunities appear to be related to male-dominated majors such as hard STEM and finance, female-dominated majors

[23] As the data collection for this chapter is the same as the previous chapter and the two research studies address the same topic (albeit with different research questions and methodologies), the literature may partly overlap.

(education, humanities and a set of social sciences) are characterized by lower work opportunities and lower salaries (Carlana & Corno 2021, Carlana 2019, Cimpan 2020, Goldin 2015, Ingellis et all 2018). Thus, persistence in gender differences in fields of study is bound to be reflected in persistence in gender inequalities in the labour market (OECD, 2021).

Understanding the origins of gender differences in college major choices is therefore crucial to evaluate the existing unbalances in the labour market, leading to key effects both at the individual and at the societal level. For individuals, the first important consequence of a wrong university choice is that it generates costs (Carlana & Corno 2021), ranging from having a lower performance at work, quitting work, or even not fitting properly into future employment. For society as a whole, wrong college major choices could cause a mismatch of talents and a loss in human capital, thus potentially reducing economic growth and development. A report by the Georgetown University Center on Education and the Workforce titled "The Economic Value of College Majors" found that choosing the wrong college major can lead to reduced earnings and a less efficient allocation of human capital in the labour market, which can have negative effects on economic growth and development. In fact, according to several studies, including those published in the Journal of Labor Economics (Arcidiacono et al., 2005), the National Bureau of Economic Research (Stinebrickner, 2014), and the report by the Georgetown University Center on Education and the Workforce (Carnevale et al., 2013), mismatched college majors can have negative effects on economic growth and development by leading to lower job satisfaction, reduced productivity, reduced earnings, and a less efficient allocation of human capital in the labour market.

Extensive literature has examined the determinants of gender disparities across different fields of study, revealing that the factors influencing college major choice are multifaceted and complex. (Bertocchi et all. 2022; Patnaik et al., 2020). Psychological and cultural factors have a great role in this context and manifest themselves both at home and at school through the negative (explicit and implicit) gender stereotypes proposed by family, teachers, and peers (Eagly et all 2019; Eccles 2011).

Carlana and Corno (2021) show that parents affect the children's field of study either by imposing direct restrictions on their choice sets or by indirectly influencing their behaviour through recommendations, while several additional contributions highlight the role of parents on children's preferences and economic/educational decisions (Doepke et al., 2019; Lizzeri and Siniscalchi, 2008; Giustinelli, 2016).

Parents' evaluations have been found to affect the association between performance-related indicators (such as teacher's ratings) and children's self-perception of ability (e.g., Frome & Eccles, 1998; Tiedemann, 2000). As regards the role of teachers, Carlana (2019) and Lavy (2008) provide evidence that teachers' stereotypes induce girls to underperform in math, develop lower self-confidence, and self-select them into less demanding high schools. As to the role of peers, Carlana and Corno (2021) and Booth et al. (2018) show that girls do not choose math to avoid interactions in male-dominated contexts (thus perpetuating gender segregation by avoiding enrolling in a field in which they know they will be rounded by the opposite sex), Shan (2020) adds that the most likely to drop out from math-related fields are the highly skilled female and Astorne-Figari and Speer (2019) point out that students switch to majors where their gender is more represented. In addition, Feld e Zolitz (2022) find evidence that students' educational choices and labour market outcomes are affected by the proportion of female peers, showing that women who are randomly assigned to a higher proportion of female peers in their studies are more likely to choose female-dominated majors (like marketing) and less likely to choose a male-dominated major (like finance).

Together with parents, teachers and peers, also role models have been shown important for college major choices (De Gioannis et all; 2023). Indeed, several studies demonstrate the positive impact of female role models on pushing female students through typical male-dominated paths, both directly by affecting their choice (Porter and Serra 2020; Breda et al. 2020) and indirectly by affecting their ability or probability to graduate (Porter and Serra 2020; Carrell et. al. 2020). Redmond et Gutke (2019) demonstrate the effectiveness of one-to-one mentorship relations in supporting females in STEM; Jethwani, et al. (2017) tests the effectiveness of live tutoring from college and graduate students together with site visits/field trips to organizations where female role models showed how they work in generating interest in female students for cybersecurity; Stoeger and al. (2016) prove the effectiveness of the E-mentoring for women in STEM with a female mentor; Merritt et al. (2021) found an increase in science identity among adolescent girls through science workshops where female professionals in STEM talked about how they became interested in STEM.

What emerges from this literature is that external factors and the context in which children grow up can significantly influence their decisions and academic achievement. Eccles' Expectancy Value Model of motivation (Eccles et al., 2002; Eccles et al., 1983) provides a useful framework for understanding the internal dynamics that can lead individuals to make biased choices or underperform (or overperform) due to external influences. The model posits that individuals' choices and performance are strongly influenced by subjective dimensions which can be shaped by the stereotypes absorbed from their environment. For instance, Schiefele et al. (2009) applied the Eccles model to investigate the influence of gender stereotypes on students' motivation and achievement in reading and writing. They found that gender stereotypes had a significant impact on students' expectancies and values in these domains, which, in turn, affected their motivation and achievement. Similarly, Denissen et al. (2018) applied the Eccles model to explore the impact of gender and ethnicity stereotypes on students' STEM career aspirations. They found that gender and ethnicity stereotypes significantly influenced students' expectancies and values for STEM careers, which, in turn, affected their career aspirations.

THE CURRENT INVESTIGATION

While most previous economic research has focused on the role of parents (Carlana and Corno 2021), teachers (Carlana 2019; and Lavy 2008), peers (Feld e Zolitz 2022; Booth et al. 2018) and role models (Porter and Serra 2020; Breda et al. 2020) in shaping students' preferences and performance, less importance has been given to the inner process that leads to the student's action (choose and/or perform). In fact, apart from a clear identification of stereotype presence, no mention is given to the determinants of choice and on how they are affected by stereotypes and their inter-relationships.

Since there is already broad theoretical and empirical support that environmental factors such as school environments and parental influences can shape (and bias) performance and college expectations, the present study hypothesizes the following:

H_1 = Students have embedded implicit stereotypes;

H_2 = The magnitude of stereotypes is different according to gender;

H_3 = Females have lower choices' determinants.

H_4 = Stereotypes affect in opposite directions -depending on gender- the choices' determinants.

H_5 = The determinants of choices have a significant impact on performance.

H_6 = The performance, in turn, affects choices.

H_7 = The determinants of choices have a significant impact on choices.

2. THEORETICAL FRAMEWORK – THE ECCLESS MODEL

The Eccles model, also known as the Eccles et al. expectancy-value theory (Eccles 2002; 1994, 1987), is a well-known theoretical framework used to explain individuals' motivation and achievement in academic and career domains. It posits that achievement-related choices and behaviors are influenced by the interaction between individuals' expectancies for success (ES) and subjective task values (STV).

The (ES) domain replies to the question *"Can I do it?"* and is shaped by personal beliefs about one's abilities. It is the self-perception of being able to achieve the goal, namely the ability's self-concept. According to the literature, this component is a strong determinant of course enrolment and occupational aspirations choices, even more, than the actual ability (Bandura, 1997; Bandura et al., 2001)

The STV, instead, replies to the question *"Why do It?"* Rokeach (1979) defined values as a set of stable and general beliefs about what is desirable, pointing out that beliefs are formed in the individual's basic psychological needs and sense of self and are influenced by societal norms. In other words, the *subjective task value* is a multidimensional construct which includes all dimensions that motivate (engage) individuals by influencing both the attractiveness of some goals and the thought that achieving those goals is something that should be done (Feather 1988, Feather 1992). It contains dimensions such as intrinsic value (I like it), attainment value (it makes me feel accomplished), utility (I value it as important for my future goals (Eccles, 1983)), costs (anxiety, fear of failure, effort) and sense of belonging.

In our model, all those variables were resumed as a subjective dimension which affects both performance and choices. Namely, we define them as the Ability Self Concept from the ES and the Cost, sense of belonging and Extrinsic and Intrinsic motivation for the STV.

All these dimensions are interacting blocks and acquire different weights depending on individuals' characteristics and surroundings. Literature shows that intrinsic motivation is a key factor both in student performance and choice (Shyr et all 2021; Orlicki 2019; Donze & Gunnes 2011; Deci, Cascio and Krussel 1975; Vallerand and Bissonnette 1992), The importance of intrinsic motivation in outcomes is also the theoretical frame that -together with empirical evidence - guides us towards model 3 in our analysis. Kirkham & Chapman (2020) found that girls showed lowered self-concept and sense of belonging in mathematical class and tend to under-aspire in their choice decision. Cost as math anxiety, nervousness and helpless are found to discourage choices towards this path by Orlicki (2019) & Dynan and Rouse (1997).

If the relative weight of each component of Eccles' model is variable and subjective, it is established that they develop together with the external stimuli derived from the environment surrounding. From the very beginning of their lives, students are exposed to external stimuli (identified with family, teachers, peers, and cultural roles) that influences students' belief and could shape their stereotypical thoughts.

Stereotypes are generalized beliefs which tend to classify people into a stigmatized category, in our case is the belief that females are suited for humanistic subjects and males for quantitative ones. Stereotypes could be explicit (conscious and externalized) or implicit. The problem with measuring explicit stereotypes is that people may not want to reveal their thoughts (social desirability), or they are not fully aware (conscious) of their feelings, so we decided to focus on implicit stereotypes[24]. We measure them by submitting students to the most widespread measure of the strength of

[24] The necessity of IAT implementation is confirmed also for our sample in which we measure both implicit and explicit Stereotypes and we found a very weak correlation between them (especially for females subsample).

stereotypical associations (De Gioannis 2022b) the Implicit Association Test (IAT) developed by social psychologists (Greenwald et al., 1998). This test, besides being widely applied by social psychologists, has also spread amongst economists, especially for research on gender (Carlana and Corno 2021, Carlana 2019, Tomasetto 2015) and cultural background (Corno 2018) inequalities. Its strength is its ability to detect biases that operate at the unconscious level (or that respondents do not want to disclose) by measuring the reaction rate. The negative impact of female stereotypes on their choices and performance has been detected in the literature (amongst others by Stoevenbelt et al. 2022; Compton et al. 2019; Spencer 1999; Steele 1997).

As theorized by Eccles' model, stereotypes could bias personal subjective feelings belonging to the expectations of success and subjective task values dimensions and, hence, indirectly educational choices and performance. Thereby, perceptions and *stereotypes* can be more relevant than interests and skills in college major choices, potentially distorting both the *expectations of success* and the *subjective task value*, and in turn the process that guides students' choices. In conclusion, in line with Eccles and Wingfield (2002)'s model, we aim at detecting how gender stereotypes affect educational choices by biasing expectations of success, perceived usefulness of the course, interest in the subject, identification with the course field and the cost of choosing one course over another. Our research fits into this context with the following innovation: starting from the variables and models and psychological literature, it proposes economic and quantitative modeling: the former (on our knowledge) on the experimental data of the analyzed geographical area.

3. DATA AND VARIABLES

DATA COLLECTION SAMPLE SELECTION

The present study used data primary data collected during a field experiment in a sample of high schools (Level 3, ISCED 2011) in two Northern districts of Italy. Data collection took place in the classroom by administering students an online questionnaire (CAWI methodology)[25] implemented on Qualtrics.

The sample consists of six high schools covering the three main types of secondary education in Italy: 2 vocational, 2 technical and 2 lyceums. Schools and classes are heterogeneous in terms of size, student characteristics and curriculum. Since the aim of the research is to investigate prevalently female behaviours, we selected mixed or female-dominated classes. Moreover, as one of the main outcomes of interest is the university enrolment intentions we have delimited the range to only the fifth grades because they are the ones where students are closest to facing this choice.

[25] Computer Assisted Web Interviewing.

The final sample size consisting of 460 students satisfies the minimum requirement (for applying the GSEM analysis) according to the general rule of thumb which requires that observation must be at least 10/20 of the variables in the model.

COVERAGE RATE, CONSISTENCY OF RECORDS AND MISSING DATA

As the data collection was administered in the classroom during regular school hours the coverage rate is around 90%: we collect information on 460 out of 536 students composing the selected classes. Undetected data only concern students who were absent that day (mainly illnesses and personal problems). This makes plausible the hypothesis that the missing students are missing completely at random (MCAR). Otherwise, filling out the questionnaire at home would probably have resulted in a lower response rate and led to biased data collection where missing values are related to the characteristics of the respondents[26] (MAR=missing at random).
Moreover, having administrated the questionnaire in class also ensures greater accuracy of the data for the following reasons: fewer misunderstandings about question compilation due to the presence (and guidance) of staff and professors, more effort and concentration from the students, little interference from the external environment, same surroundings, and device for all respondents in both pre and post questionnaires. All these points (especially the third) besides being very important for the whole questionnaire acquire even more importance in the IAT exercise. In fact, more than 97% of students carried out the IAT correctly, only some detections were declared missing because they did not meet reliability criteria. Missingness is imposed by the software and arises when it detects that the test was done without diligence and effort. Namely following the procedure first reported by Greenwald et. al., 2003 and resumed by Lane et al., 2012 (p. 92), trials are classified as missing trials if they are too slow (over 10,000 ms) and or if participants provide more than 10% of the stimuli faster than 300 ms.

VARIABLES DESCRIPTION

DEPENDENT VARIABLES

EcoMath, STEM, Hum = three continuous variables, ranging from 1 *"Extremely Unlikely"* to 5 "*Extremely Likely*", reflecting students' likelihood to enrol in each specific university path. Students were asked about their intentions after completing high school between 15 options: enrolment in 14 university fields and going to work. Then, where created the Macro-group on university choices

[26] It is reasonable to assume that the students who do not return the questionnaire are: the least diligent, the most careless, those with poorer family circumstances (no device), etc.

follows MIUR[27] aggregation, further details on group definition and individual university fields could be found in Annex N° 1 and in table A1. The aggregation procedure was done by assigning to the macro group the highest value achieved among its sub-fields when taken individually. As we are dealing with enrolment intention, we judged that the most appropriate way to measure the likelihood of going to a particular area was to take the highest reached in its options as the most probable path. to give an example: if within EcoMath we have the following scores Mathematics (1) Economics and marketing (3) Economics and finance (5) Statistics (2), the probability of choosing Ecomath will be high (5 = extremely likely) because of the propensity to take a course in economics and finance. As the final students' choices should be one, we think it is less correct, to average the scores obtained in each option[28], since it is irrelevant if the student does not want to enrol in math or statistics because he will choose EcoMath anyway due to the high economic propensity. Nor was the min-max normalization procedure included because GSEM allows the study of multiple independent variables simultaneously: we suggest this latter in the case you want to consider in your model one choice' field at a time, as the dependent variable created is an absolute rather than a relative[29] value.

Finally, where censored (set to zero) the enrolment propensity for those students who declared they are strictly motivated[30] to go to work (rather than choosing university) since their inclination toward a determined field of study is unlikely to result in actual enrolment.

Between the four macro groups created with this procedure, we selected the three groups concerning economics and math, STEM and humanistic fields; "health and medicine" was not considered in our analysis because it was distant from the research interests and covariates collected.

Ability Gap = is a continuous variable (ranging from -5.5 to 2 in our sample) which took negative values if the students perform better in humanistic subjects respect the mathematical ones and positive values for the opposite cases. The variable was created by averaging students' reported grades at the end of the previous year in economics[31] and maths (Cronbach's alpha $\alpha = 0.74$) and in Italian and history (Cronbach's alpha $\alpha = 0.81$) and subtracting the latter from the former. This variable was created with the aim to observe the relative ability rather than the absolute one: not

[27] The Italian Ministry of Education, University and Research.
[28] Following the example before the EcoMath score obtained by applying the averaging procedure should be 2,75.
[29] Rate Ecomath 3 (Neither likely nor unlikely) acquires an intensely different result if the score in all other choices is 1 (extremely likely) or 5 (extremely unlikely).
[30] Strictly work motivation arises when students give the highest rating to the option of going to work with no equal among the other 14 choices.
[31] Not all students have economics as a subject and, in those cases, only the maths mark is taken.

students' ability in the subjects, but whether they are better at humanities or mathematics. This variable enters the model both as dependent and independent in explaining students' choices.

INDEPENDENT VARIABLES

Stereotype = Continuous variable (ranging from -0.81 to 1.26 in our sample). Stereotypes were detected through the Implicit Association Test (IAT) which measures implicit stereotypes: in our case automatic beliefs about the association between gender and a given field of education (De Gioannis, 2022a; Martin & Dinella, 2001). Stereotypes are detected through differences in students' reaction time in categorizing words concerning gender and field of study: positive values indicate a positive mental association between female and humanistic field and male and economic|mathematic fields, neutral values reflect the absence of stereotypes and negative values the opposite mental association (male - humanistic and female - economic|mathematic fields). Further details about how IAT works, the stimuli we used and the procedure we implemented can be found in Annex 1 of Addabbo, Strozzi, Tasselli 2022 which is the first essay of this thesis.

Intrinsic = Continuous variable (ranging from -1.4 to 1.88 in our sample) indicating how much students like maths and economics subjects. The higher the value the more students like the subject. This variable originated from two different questions which were standardized to homologate the scale and averaged (Cronbach's alpha α[32] = 0.76).

Extrinsic = Continuous variable ranging from 1 *"not important"* to 5 "*extremely important"* which corresponds to how important students perceive the economy to be for their professional, academic and general future.

The variable originated from three different questions that were averaged (Cronbach's alpha α = 0.87).

Costs = Continuous variable ranging from 1 *"low perceived cost"* to 4 *"high perceived cost"* which averaged replies to the following question extrapolated from a PISA' battery: I am often worried that maths lessons could be difficult for me; I am very nervous when I must do a maths assignment.; I am very nervous when I must solve a maths problem; I am afraid that I will get low grades in mathematics. (Cronbach's alpha α = 0.86).

Ability S.C. (ability self-concept) = Continuous variable ranging from 1 *"low ability self-perception"* to 4 *"high ability self-perception*". The variable was created by using 4 variables

[32] Cronbach's alpha (Cronbach, 1951), also known as coefficient alpha, is a measure of reliability, particularly the reliability of the internal consistency or item. measure of reliability, particularly the reliability of the internal consistency or interrelatedness of a scale or test (e.g., a questionnaire). Internal consistency refers to the extent to which all items in a scale or test contribute positively to the measurement of the same construct. Cronbach's alpha typically ranges from 0 to 1. Values closer to 1.0 indicate a greater internal consistency of the variables in the scale. In other words, higher Cronbach's alpha values show greater scale reliability. A frequently cited acceptable range of Cronbach's alpha is a value of 0.70 or above. This is derived from the work of Nunnally (1978).

extrapolated from the PISA' battery used for the Cost indicator. Variables are statements about how students feel good in maths (Cronbach's alpha α = 0.84).

Belonging (sense of belonging) = Continuous variable ranging from 1 *"low sense of belonging to the economic community"* to 8 *"high sense of belonging to the economic community"* The variables included in the index explore the student's feeling by asking him/her to imagine him/herself in an economics and finance class). The indicators of this variable were inspired by the reduced scale of Good et al. 2012. Some variables were reversed to homologate the direction and then averaged (Cronbach's alpha α = 0.82).

We also tried to aggregate the composite indicators (Intrinsic, Extrinsic, Cost Ability S.C. and Belonging) through separated Structural Equation Modelling in which they were defined as latent variables and computed as a factor score using the list of variables that respectively concerned them. The results obtained were very similar[33] to those from the mediation procedure, so we preferred the latter.

4. DESCRIPTIVE STATISTICS

Table 1 shows general information about the sample we are dealing with, while the main variables of the model are approached from Table 2 onward. Observation across variables are unbalanced because of some missing data. 65% of the sample are female and the average age is around 19 years (there is little variation in age as students belong to the same academic school cohort). 22% of them have repeated at least one class, and they are manly born in the Emilia Romagna region (only 7% were born in another region of Italy and 8% abroad). The average years of education of their parents are approximately 12 for both parents (slightly higher for mothers). Higher percentage of fathers undertook quantitative studies (Scientific lyceum or a stem|economic degree). The level of mothers' unemployment is significantly higher than the one of fathers and most importantly no unemployed fathers were declared as *housemakers*, while almost all unemployed mothers were. Finally, concerning the school involved, nearly to $\frac{1}{2}$ of the students came from technical institutes, $\frac{1}{3}$ from lyceum and the remaining $\frac{1}{5}$ from professional institutes. The class size varies from a minimum of 13 to a maximum of 26 students per class with an average of 21 students per class.

Table 1 – Descriptive Characteristics of the sample

Variable	Obs	Mean	Std. Dev.	Min	Max
PANEL A – STUDENTS' CHARACTERISTICS					
Female	460	.65	.48	0	1

[33] The correlation between the indicators derived with the two different methodologies was around 0.99 for all variables.

Age	464	19.22	.61	17	22
Repeat	463	.22	.42	0	1
Born in ER	464	.84	.37	0	1
Born in IT	464	.07	.26	0	1
Born abroad	464	.09	.28	0	1
PANEL B – FAMILY BACKGROUND					
Mother_schooling	429	12.52	3.43	8	21
Father_schooling	423	11.88	3.49	8	21
Mother_quantitative	429	.07	.25	0	1
Father_quantitative	423	.08	.27	0	1
Mother_notemployed	465	.18	.38	0	1
Father_notemployed	465	.05	.21	0	1
PANEL C – SCHOOL ENVIROMENT					
Lyceum	465	.33	.47	0	1
Technical	465	.47	.5	0	1
Vocational	465	.2	.4	0	1
Class Size	465	20.98	3.64	13	26

Table 2 – Bivariate statistics: gender comparison of major choices determinants and choices

	(1) FEMALE			(2) MALE			(3) T-TEST ($\bar{x}_f - \bar{x}_m$)	
	count	**mean**	**sd**	**count**	**mean**	**sd**	**b**	**t**
PANEL A – Choices determinants								
Stereotype	295	0.40	0.41	153	0.14	0.40	0.25***	(6.25)
Intrinsic	297	-0.08	0.93	162	0.08	0.95	-0.16	(-1.74)
Extrinsic	297	2.99	1.21	162	3.19	1.02	-0.20	(-1.92)
Costs	297	2.71	0.84	162	2.29	0.77	0.43***	(5.46)
S_belonging	297	4.37	1.26	162	4.87	1.20	-0.50***	(-4.17)
Abilitysc	297	2.37	0.74	162	2.51	0.76	-0.15*	(-2.00)
Abilitygap	297	-0.18	1.02	162	-0.35	1.06	0.17	(1.68)
Mark_hum	297	7.42	0.90	162	7.04	0.88	0.37***	(4.33)
Mark_ecomath	297	7.24	1.20	162	6.69	1.20	0.55***	(4.66)
PANEL B[1] – choices intentions (work and majors)								
Work_raw	297	3.01	1.41	162	3.22	1.40	-0.21	(-1.50)
Ecomath_raw	297	2.51	1.37	162	2.74	1.39	-0.23	(-1.72)
Stem_raw	297	2.10	1.25	162	2.22	1.33	-0.11	(-0.88)
Hum_raw	297	3.22	1.42	162	2.45	1.36	0.76***	(5.67)
PANEL C – choices intentions (work and majors; major are truncated at strictly high work probability)								
Worker [2]	297	0.22	0.42	162	0.37	0.48	-0.15**	(-3.29)
Ecomath[3]	231	2.65	1.44	102	3.11	1.44	-0.46**	(-2.70)
Stem[3]	231	2.20	1.32	102	2.53	1.43	-0.33*	(-1.99)
Hum[3]	231	3.45	1.43	102	2.86	1.39	0.59***	(3.52)
N	**297**			**162**			**459**	

Notes:

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[1] The average likelihood of taking the choice (going to work or enrolling in a specific field).

[2] Strictly high work probability.

[3] The average likelihood of enrolling in a specific field given null strictly high work probability. Workers were declared as missing only in descriptives statistics to show the number of observations which intended to go to university and their actual average, but in the model were declared at zero and censored.

Table 2 presents the descriptive statistics for the variables related to choice determinants and choice intention by gender. The differences between the female and male averages were calculated by subtracting the female average ($\bar{x}_f$) from the male average ($\bar{x}_m$), where a positive sign indicates that women's characteristics are higher than men's, and a negative sign indicates the opposite.

Regarding IAT averages, both males and females suffer from implicit stereotyping, but for females, this belief is more than twice as high as for males and the t-test confirms that difference at all significance levels ($p< 0.01$). As reported in the first column (count), the IAT score is the only variable in which we have missing values[34], however, they are less than 0,7% for the female population and around 5% for males.

Concerning choices determinants do not appear to be a statistically significant difference by gender in math or economics enjoyment nor in extrinsic motivation (even if for both indicators male averages are higher).

On the other hand, male students have a deeper sense of belonging and have lower psychological costs in addressing subjects with higher math content and the difference is significant at 1% (p-value < 0.01) .

Finally, it is interesting to notice that even if females outperform men in math and economics grades (7.24 vs 6.69), their overconfidence in this field is lower (ability self-perception). The difference between the performances in the humanistic and the quantitative subject (ability gap) is not significant since female students outperform male students also in humanities (7.42 vs 7.04).

Panel B and panel C both report enrolment (or work) intentions, the difference is that the former report the raw scores which do not account for students strongly intent on working after high school, meanwhile the latter applicate this correction by censoring the *Ecomath, Stem, Hum* variables when *Worker* is equal to one. *Worker* is a dummy variable equal to one if the student rate the option "*going to work*" as the strictly[35] highest among all other choice options.

Even if in both panels the gender gaps have the same sign, we can see that in Panel B all coefficients are smoothed by the influence of work intention; meanwhile in Panel C gender differences in all choices (including that of working) became significant.

Table 3 – Bivariate correlation between stereotypes, determinants and major choice by gender

PANEL A - Female											
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Stereotype	1.00										
(2) Intrinsic	-0.16***	1.00									
(3) Extrinsic	-0.28***	0.42***	1.00								
(4) Cost	-0.02	-0.55***	-0.10*	1.00							
(5) Belonging	-0.18***	0.38***	0.43***	-0.20***	1.00						

[34] Missingness is imposed for those results which are not reliable.

[35] University propensity is still observed if student rate with the same probability "going to work" and at least one other choice.

(6) Abilitysc	-0.06	0.75***	0.28***	-0.70***	0.29***	1.00					
(7) Abilitygap	-0.14**	0.48***	0.24***	-0.34***	0.20***	0.45***	1.00				
(8) Worker	0.05	-0.14**	0.02	0.07	-0.04	-0.17***	0.03	1.00			
(9) Ecomath	-0.28***	0.29***	0.53***	-0.01	0.52***	0.17***	0.13**	NA	1.00		
(10) Stem	-0.11*	0.26***	0.15**	-0.14**	0.09	0.16**	0.26***	NA	0.04	1.00	
(11) Hum	0.19***	-0.44***	-0.36***	0.25***	-0.32***	-0.36***	-0.30***	NA	-0.33***	-0.36***	1.00

PANEL B - Male											
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Stereotype	1.00										
(2) Intrinsic	0.13*	1.00									
(3) Extrinsic	0.26***	0.27***	1.00								
(4) Costs	-0.15*	-0.49***	-0.11	1.00							
(5) Belonging	0.17**	0.18**	0.37***	-0.23***	1.00						
(6) Abilitysc	0.16**	0.81***	0.24***	-0.63***	0.23***	1.00					
(7) Abilitygap	0.13*	0.45***	0.25***	-0.37***	0.27***	0.53***	1.00				
(8) Worker	-0.06	-0.10	-0.01	0.11	-0.06	-0.11	-0.06	1.00			
(9) Ecomath	0.23**	0.31***	0.50***	-0.20**	0.29***	0.34***	0.38***	NA	1.00		
(10) Stem	-0.07	0.31***	0.12	-0.05	-0.18*	0.31***	0.03	NA	0.05	1.00	
(11) Hum	-0.17*	-0.20**	-0.36***	0.08	-0.08	-0.21**	-0.36***	NA	-0.16	-0.14	1.00

Notes:
**** p<0.01, ** p<0.05, * p<0.1*
Cells report Pearson correlation

Table 3 reports the Pearson coefficient (r) for the bivariate correlation between stereotypes (1), choices determinants (from 2 to 7) and enrolment intention (from 8 to 11). Correlations were computed separately by gender for the inverse impact that stereotypes cause according to gender. In fact, we can see that column one is significantly and negatively correlated with almost all female indicators, while this correlation for men is positive. Females who suffer from high stereotypes tend to have a lower sense of belonging, extrinsic and intrinsic motivation and underperform in quantitative subjects. As regards university choices, they prefer humanities and avoid stem and economics majors: in particular, the strongest correlation for female stereotypes regards exactly the enrolment in economics intention. On the opposite, stereotypes boost male students' sense of belonging, intrinsic motivations, ability self-concept and observed ability with the greater effect on their perceived future utilities and their enrolment intentions in economics and mathematics universities.

In both panel workers' variables have any significative correlation because is related to another type of choice (working respect studying rather than fields inclination or preferences). On the other side, all other variables have a strong and significant correlation with the expected sign (negative only for cost and Humanities): a sign of internal consistency among the indicators chosen and between them and university propensities.

5. METHODOLOGY

We employed a generalized structural equation model (GSEM) to analyse the impact of choice determinants on the student's enrolment intentions and the intra-relationships between determinants and stereotypes. GSEM is a particular variant of the structural equation model (SEM) which combines multiple regression analyses into a single model to simultaneously predict different outcomes and complex relationships between independent and dependent variables. We opt for GSEM rather than SEM because the latter permits only to include continuous variables with a normal distribution and does not permit multilevel model and interaction terms the former is more in line with our empirical needs since it allows to include important variables as binary variables and other kinds of generalized responses (StataCorp 2021), to control for the interaction between gender and stereotypes and for the fact that students are nested within schools. Two empirical assumptions which are of crucial importance because:

i) stereotypes have an inverse impact on determinants depending on gender

ii) the school attended has a nonnegligible relationship with university choice.

Moreover, the use of GSEM, also allowed us to deal with censored dependent variables. In fact (as explained in the section dedicated to variables definition) to better approximate the reality we declared as university choices unobserved for those students who are highly intentioned not to proceed with university at the end of their high school but rather to go to work.

The conceptual framework behind our model is visualized in Figures 1, 2 and 3. All models were inspired by Eccles and Wingfield's model (2002) and are funded in the theoretical framework presented in section 2.

Finally, each model included a school-level random intercept: a latent variable at the school level that is constant within schools and varies across schools to account for the fact that students nested within schools are more likely to make the same choices than students from different schools. This assumption becomes even stronger in the Italian context as the choice of high school (although not binding) is already oriented/orienting of/the post-diploma. This led to the model becoming a two-level (nested-effects) model which better approximates reality, as students are nested between schools and choices are strongly dependent on schools. The reasons for this further correction are confirmed by the fact that in all three models, the Likelihood-ratio test after estimation (Greene 2018, 554 – 555) rejects at any reasonable level that the model without the latent correction adequately accounts for school clusters. Moreover, the Akaike information criterion (AIC) (Akaike 1987) and Schwarz's Bayesian Information Criterion (BIC) (Raftery 1993) are always both lower for the models with the random intercept (see section 4 dedicated to results and table X in Appendix).

MODEL ONE

Model 1, displayed in figure one, implies that stereotypes affect subjective perceptions of individuals (intrinsic and extrinsic motivations; the perceived cost associated with the subject, the sense of belonging to the economics and mathematics field and the ability self-concept). The interaction between gender and stereotypes is essential to have significant (and unbiased) coefficients; in fact, stereotypes affect the subjective perception in an opposite direction according to gender. This is because suffering from stereotypes means associating females with humanistic subjects and males with economic and mathematic ones, so stereotypes affect self-confidence and beliefs in reverse ways depending on one's gender. Moving forward to the right in the interpretation of the figure, we see how all subjective indicators influence ability and fields' choices; the latter are influenced both directly and indirectly by the ability gap.

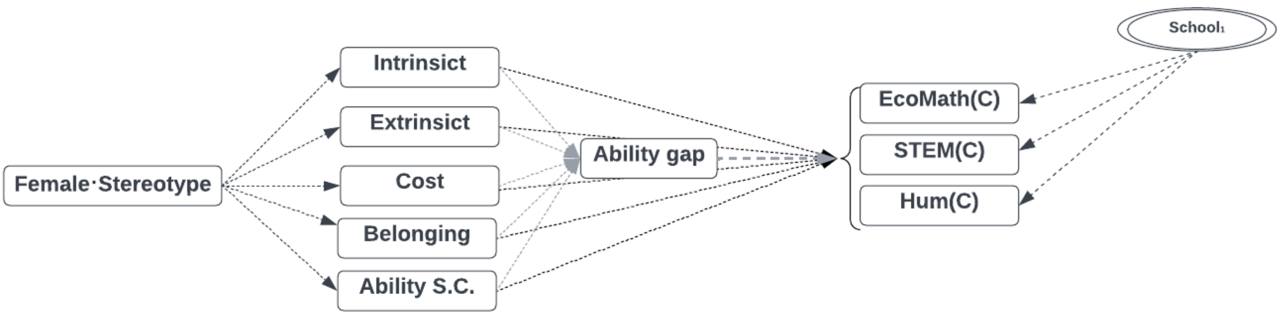


Figure 1 – Impact of choices determinants on ability gap and on choices (both directly and indirectly through ability gap)

Equation 1 shows the system of equations associated with the theoretical model displayed in figure 1.

$$\begin{cases} I = \alpha_0 + \beta_1 F + \beta_2 S + \beta_3 F \cdot S + \varepsilon_1 \\ E = \alpha_1 + \beta_4 F + \beta_5 S + \beta_6 F \cdot S + \varepsilon_2 \\ C = \alpha_2 + \beta_7 F + \beta_8 S + \beta_9 F \cdot S + \varepsilon_3 \\ B = \alpha_3 + \beta_{10} F + \beta_{11} S + \beta_{12} F \cdot S + \varepsilon_4 \\ AS = \alpha_4 + \beta_{13} F + \beta_{14} S + \beta_{15} F \cdot S + \varepsilon_5 \\ AG = \alpha_5 + \beta_{16} I + \beta_{17} E + \beta_{18} C + \beta_{19} B + \beta_{20} AS + \varepsilon_6 \\ ECOM^* = \alpha_6 + \beta_{21} I + \beta_{22} E + \beta_{23} C + \beta_{24} B + \beta_{25} AS + \beta_{26} AG + \gamma_1 M_{1,S} + \varepsilon_6 \\ STEM^* = \alpha_7 + \beta_{27} I + \beta_{28} E + \beta_{29} C + \beta_{30} B + \beta_{31} AS + \beta_{32} AG + \gamma_2 M_{1,S} + \varepsilon_7 \\ HUM^* = \alpha_8 + \beta_{33} I + \beta_{34} E + \beta_{35} C + \beta_{36} B + \beta_{35} AS + \beta_{37} AG + \gamma_3 M_{1,S} + \varepsilon_8 \end{cases} \quad (1)$$

$$ECOM = \begin{cases} ECOM^* \text{ if } ECOM^* > 0 \\ 0 \text{ if } ECOM^* \leq 0 \end{cases} \quad (2a)$$

$$STEM = \begin{cases} STEM^* \text{ if } STEM^* > 0 \\ 0 \text{ if } STEM^* \leq 0 \end{cases} \quad (2b)$$

$$HUM = \begin{cases} HUM^* \text{ if } HUM^* > 0 \\ 0 \text{ if } HUM^* \leq 0 \end{cases} \quad (2c)$$

$$(\varepsilon_1; \varepsilon_2; \varepsilon_3; \varepsilon_4, \varepsilon_5; \varepsilon_6; \varepsilon_7; \varepsilon_8) \sim i.i.d. \text{ with } E(\varepsilon_j) = 0 \text{ for all } j$$
$$Cov(\varepsilon_j, \varepsilon_k) = 0 \text{ for j not equal to k, j,k } = 1, \ldots, 8.$$

$$Cov(\varepsilon_1, I) = 0 \qquad Cov(\varepsilon_4, B) = 0 \qquad Cov(\varepsilon_7, ECOM^*) = 0$$
$$Cov(\varepsilon_2, E) = 0 \qquad Cov(\varepsilon_5, AS) = 0 \qquad Cov(\varepsilon_8, STEM^*) = 0$$
$$Cov(\varepsilon_3, C) = 0 \qquad Cov(\varepsilon_6, AG) = 0 \qquad Cov(\varepsilon_9, HUM^*) = 0$$

Where i.i.d. means that observations are independent and identically distributed.
Where F refers to the female dichotomous variable and S is the stereotype continuous one.
Where I, E, C, B, AS, AG correspond respectively to the indicator for Intrinsic, Cost, Ability Self Concept, and Ability Gap. M_1 correspond to a latent variable random effect which accounts for the fact that students are nested within schools (s = 1, … ,6). In the latent variable M, gamma was constrained to be equal to one ($\gamma = 1$) .
Moreover, as reported in equations 2a; 2b; 2c for each of the choice's intentions (Economics and maths (2.a ECOM), 2.b STEM and 2.c Humanities (HUM)) we got unobserved values (ranked as 0 lower limits) for students who declare they will work. This corresponds to the standard Tobit model (Tobin 1958) in which we have a dependent variable y that is left-censored at zero.
In each eq. 2, the actual value for the dependent variable *choice* (in turn ecom, STEM, hum) is observed if the latent variable *choice** is above 0 (lower limit), meanwhile, the zero' limit is taken for the censored observations.

MODEL TWO

Model two (Figure 2 and equation 3), follows the theory and the assumptions of model one. What changes is that, in this model, subjective values have been separated into their two main subsections: Subjective task values and expectation of success. This implies that the second level has two distinct dimensions: the STV (intrinsic, Extrinsic, Cost and Belonging) which directly affect choices and the ES in which ability self-concept affects the ability gap and through (and jointly) this affects the choice. The choice's form and assumption remain the same as in model 1.

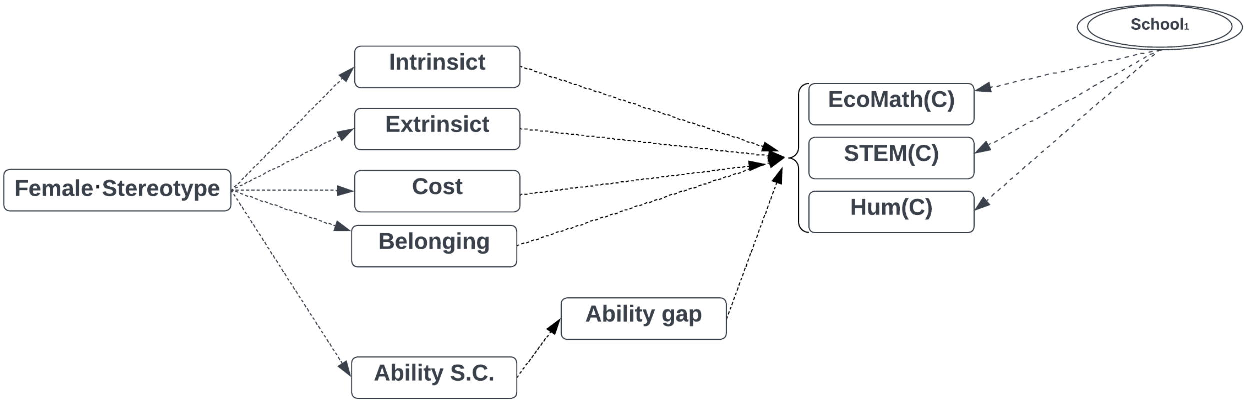


Figure 2 – Impact of choices determinants on choices (the partition between the Expectation of Success and Subjective Task Value)

$$\begin{cases} I = \alpha_0 + \beta_1 F + \beta_2 S + \beta_3 F \cdot S + \varepsilon_1 \\ E = \alpha_1 + \beta_4 F + \beta_5 S + \beta_6 F \cdot S + \varepsilon_2 \\ C = \alpha_2 + \beta_7 F + \beta_8 S + \beta_9 F \cdot S + \varepsilon_3 \\ B = \alpha_3 + \beta_{10} F + \beta_{11} S + \beta_{12} F \cdot S + \varepsilon_4 \\ AS = \alpha_4 + \beta_{13} F + \beta_{14} S + \beta_{15} F \cdot S + \varepsilon_5 \\ AG = \alpha_5 + \beta_{16} AS + \varepsilon_6 \\ ECOM^* = \alpha_6 + \beta_{17} I + \beta_{18} E + \beta_{19} C + \beta_{20} B + \beta_{21} AG + \gamma_1 M_{1,S} + \varepsilon_6 \\ STEM^* = \alpha_7 + \beta_{22} I + \beta_{23} E + \beta_{24} C + \beta_{25} B + \beta_{26} AG + \gamma_2 M_{1,S} + \varepsilon_7 \\ HUM^* = \alpha_8 + \beta_{27} I + \beta_{28} E + \beta_{29} C + \beta_{30} B + \beta_{31} AG + \gamma_3 M_{1,S} + \varepsilon_8 \end{cases} \tag{3}$$

$$ECOM = \begin{cases} ECOM^* \text{ if } ECOM^* > 0 \\ 0 \text{ if } ECOM^* \leq 0 \end{cases} \tag{4a}$$

$$STEM = \begin{cases} STEM^* \text{ if } STEM^* > 0 \\ 0 \text{ if } STEM^* \leq 0 \end{cases} \tag{4b}$$

$$HUM = \begin{cases} HUM^* \text{ if } HUM^* > 0 \\ 0 \text{ if } HUM^* \leq 0 \end{cases} \tag{4c}$$

$(\varepsilon_1; \varepsilon_2; \varepsilon_3; \varepsilon_4, \varepsilon_5; \varepsilon_6; \varepsilon_7; \varepsilon_8) \sim i.i.d.$ with $E(\varepsilon_j) = 0$ for all j
$Cov(\varepsilon_j, \varepsilon_k) = 0$ for j not equal to k, j,k =1,. . . ,8.

$Cov(\varepsilon_1, I) = 0$	$Cov(\varepsilon_4, B) = 0$	$Cov(\varepsilon_7, ECOM^*) = 0$
$Cov(\varepsilon_2, E) = 0$	$Cov(\varepsilon_5, AS) = 0$	$Cov(\varepsilon_8, STEM^*) = 0$
$Cov(\varepsilon_3, C) = 0$	$Cov(\varepsilon_6, AG) = 0$	$Cov(\varepsilon_9, HUM^*) = 0$

MODEL THREE

Model 3 (displayed in figure 3 and equation 4) is a hybrid of the first two which combines theoretical and empirical justification.

In this model (as in model 2) the partition between STV and ES remains, but moreover, the impact of intrinsic motivation on the ability gap is reintroduced as in model 1.

The draft of this additional path has empirical bases: the results of model 1 show a significant impact of intrinsic motivation on the ability gap. Moreover, further confirmation comes from the improvement of the goodness of fit (model 3 is the one with the lowest Comparative Fit Index and Akaike information criterion.

In addition, this path also has strong theoretical support; in fact, there is much research confirming the relevance of intrinsic motivation on learning outcomes (Froiland & Worrell 2016; Augustyniak et all. 2016; Heyman et all, 1992; Adelman & Taylor, 1986; Adelman, 1978).

The assumptions of models 1 and 2 together with the choice specification are also reiterated for this model.

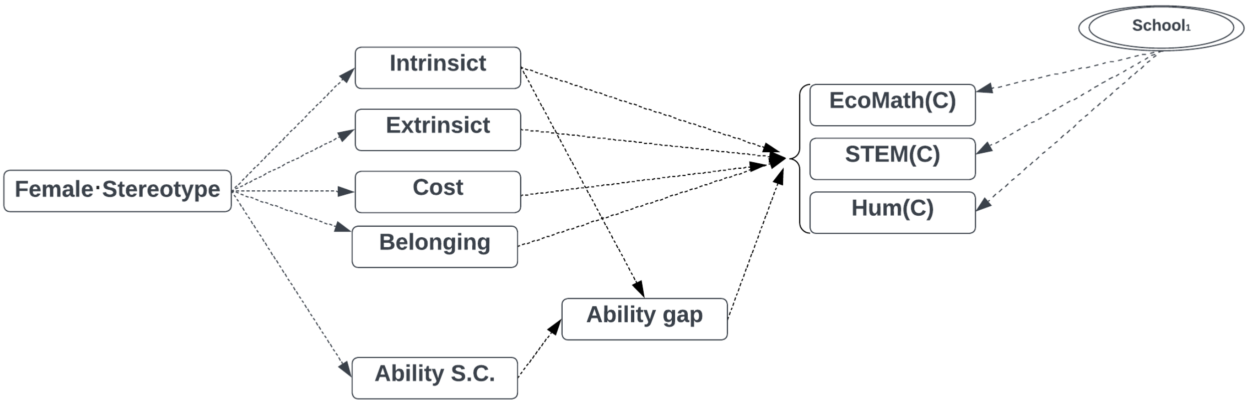


Figure 3 – Impact of choices determinants on choices (the partition between the Expectation of Success and Subjective Task Value) + Intrinsic theory.

$$\begin{cases} I = \alpha_0 + \beta_1 F + \beta_2 S + \beta_3 F \cdot S + \varepsilon_1 \\ E = \alpha_1 + \beta_4 F + \beta_5 S + \beta_6 F \cdot S + \varepsilon_2 \\ C = \alpha_2 + \beta_7 F + \beta_8 S + \beta_9 F \cdot S + \varepsilon_3 \\ B = \alpha_3 + \beta_{10} F + \beta_{11} S + \beta_{12} F \cdot S + \varepsilon_4 \\ AS = \alpha_4 + \beta_{13} F + \beta_{14} S + \beta_{15} F \cdot S + \varepsilon_5 \\ AG = \alpha_5 + \beta_{16} I + \beta_{17} AS + \varepsilon_6 \\ ECOM^* = \alpha_6 + \beta_{18} I + \beta_{19} E + \beta_{20} C + \beta_{21} B + \beta_{22} AG + \gamma_1 M_{1,S} + \varepsilon_6 \\ STEM^* = \alpha_7 + \beta_{23} I + \beta_{24} E + \beta_{25} C + \beta_{26} B + \beta_{27} AG + \gamma_2 M_{1,S} + \varepsilon_7 \\ HUM^* = \alpha_8 + \beta_{28} I + \beta_{29} E + \beta_{30} C + \beta_{31} B + \beta_{32} AG + \gamma_3 M_{1,S} + \varepsilon_8 \end{cases} \tag{4}$$

$$ECOM = \begin{cases} ECOM^* \text{ if } ECOM^* > 0 \\ 0 \text{ if } ECOM^* \leq 0 \end{cases} \tag{2a}$$

$$STEM = \begin{cases} STEM^* \text{ if } STEM^* > 0 \\ 0 \text{ if } STEM^* \leq 0 \end{cases} \tag{2b}$$

$$HUM = \begin{cases} HUM^* \text{ if } HUM^* > 0 \\ 0 \text{ if } HUM^* \leq 0 \end{cases} \tag{2c}$$

$(\varepsilon_1; \varepsilon_2; \varepsilon_3; \varepsilon_4, \varepsilon_5; \varepsilon_6; \varepsilon_7; \varepsilon_8) \sim i.i.d.$ with $E(\varepsilon_j) = 0$ for all j
$Cov(\varepsilon_j, \varepsilon_k) = 0$ for j not equal to k, j,k =1,. . . ,8.

$Cov(\varepsilon_1, I) = 0$ $\quad Cov(\varepsilon_4, B) = 0$ $\quad Cov(\varepsilon_7, ECOM^*) = 0$
$Cov(\varepsilon_2, E) = 0$ $\quad Cov(\varepsilon_5, AS) = 0$ $\quad Cov(\varepsilon_8, STEM^*) = 0$
$Cov(\varepsilon_3, C) = 0$ $\quad Cov(\varepsilon_6, AG) = 0$ $\quad Cov(\varepsilon_9, HUM^*) = 0$

The following paragraph reports the results of equations (1); (3); (5).

Since GSEM does not allow for the direct calculation of standardized results, all variables were divided for their standard deviation in order to homogenize their unit of measurement and make the coefficients (β) comparable across variables.

As GSEM doesn't allow to compute the CFI (Comparative Fit Index) or other common indicators of model fit (mean square error of approximation, the model's Chi-square) we judge the goodness of fit by using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as literature usually does (Fong and Kremer 2020; Perez et al. 2014).

6. RESULTS

Table 4 – Estimates from the Generalized structural equation modelling

	MODEL (1)	MODEL (2)	MODEL (3)
PANEL A – FIRST LEVEL: Influence of stereotype on choice determinants.			
Intrinsic			
Female	0.03	0.03	0.03
	(0.28)	(0.28)	(0.28)
Stereotype	0.14	0.14	0.14
	(1.67)	(1.67)	(1.67)
female·Stereotype	-0.30**	-0.30**	-0.30**
	(-2.93)	(-2.93)	(-2.93)
Constant	0.04	0.04	0.04
	(0.49)	(0.49)	(0.49)
Extrinsic			
Female	0.19	0.19	0.19
	(1.65)	(1.65)	(1.65)
Stereotype	0.25**	0.25**	0.25**
	(2.94)	(2.94)	(2.94)
female·Stereotype	-0.55***	-0.55***	-0.55***
	(-5.43)	(-5.43)	(-5.43)
Constant	2.69***	2.69***	2.69***
	(32.42)	(32.42)	(32.42)
Costs			
Female	0.47***	0.47***	0.47***
	(4.06)	(4.06)	(4.06)
Stereotype	-0.15	-0.15	-0.15
	(-1.78)	(-1.78)	(-1.78)
female·Stereotype	0.13	0.13	0.13
	(1.27)	(1.27)	(1.27)

Constant	2.78***	2.78***	2.78***
	(33.23)	(33.23)	(33.23)
Belonging			
Female	-0.16	-0.16	-0.16
	(-1.46)	(-1.46)	(-1.46)
Stereotype	0.17*	0.17*	0.17*
	(2.04)	(2.04)	(2.04)
female·Stereotype	-0.35***	-0.35***	-0.35***
	(-3.49)	(-3.49)	(-3.49)
Constant	3.78***	3.78***	3.78***
	(45.66)	(45.66)	(45.66)
Ability S. C.			
Female	-0.08	-0.08	-0.08
	(-0.67)	(-0.67)	(-0.67)
Stereotype	0.18*	0.18*	0.18*
	(2.10)	(2.10)	(2.10)
female·Stereotype	-0.24*	-0.24*	-0.24*
	(-2.26)	(-2.26)	(-2.26)
Constant	3.30***	3.30***	3.30***
	(38.51)	(38.51)	(38.51)
PANEL B – SECOND LEVEL: Influence of choice determinants on Ability.			
Ability Gap			
Intrinsic	0.22**		0.26***
	(3.24)		(3.96)
Extrinsic	0.07		
	(1.50)		
Costs	-0.01		
	(-0.16)		
Belonging	0.02		
	(0.49)		
Ability S. C.	0.27***	0.47***	0.27***
	(3.70)	(11.38)	(4.24)
Constant	-1.34***	-1.75***	-1.11***
	(-3.46)	(-12.51)	(-5.20)
PANEL C – THIRD LEVEL: Influence of choice determinants on enrollment intention.			
Hum			
Intrinsic	-0.01	0.09	0.09
	(-0.06)	(0.87)	(0.87)
Extrinsic	-0.26**	-0.25**	-0.25**
	(-2.81)	(-2.79)	(-2.79)
Costs	0.10	0.03	0.03
	(0.92)	(0.32)	(0.32)
Belonging	-0.05	-0.04	-0.04
	(-0.52)	(-0.49)	(-0.49)
Ability S. C.	0.19		
	(1.33)		
Ability Gap	-0.10	-0.08	-0.08
	(-1.12)	(-0.88)	(-0.88)
M1[school]	1.00	1.00	1.00
	(.)	(.)	(.)
Constant	1.23	2.04***	2.04***

	(1.55)	(4.05)	(4.05)
Ecomath			
Intrinsic	0.15	0.25**	0.25**
	(1.32)	(2.68)	(2.68)
Extrinsic	0.25**	0.25**	0.25**
	(3.09)	(3.09)	(3.09)
Costs	0.10	0.03	0.03
	(1.06)	(0.39)	(0.39)
Belonging	0.24**	0.24**	0.24**
	(3.01)	(3.04)	(3.04)
Ability S. C.	0.19		
	(1.50)		
Ability Gap	0.02	0.05	0.05
	(0.30)	(0.59)	(0.59)
M1[school]	1.00	1.00	1.00
	(.)	(.)	(.)
Constant	-1.35	-0.53	-0.53
	(-1.90)	(-1.16)	(-1.16)
Stem			
Intrinsic	0.29*	0.36***	0.36***
	(2.57)	(3.94)	(3.94)
Extrinsic	0.03	0.03	0.03
	(0.39)	(0.40)	(0.40)
Costs	0.02	-0.03	-0.03
	(0.26)	(-0.32)	(-0.32)
Belonging	-0.06	-0.06	-0.06
	(-0.80)	(-0.77)	(-0.77)
Ability S. C.	0.14		
	(1.14)		
Ability Gap	0.05	0.06	0.06
	(0.59)	(0.82)	(0.82)
M1[school]	1.00	1.00	1.00
	(.)	(.)	(.)
Constant	0.54	1.14*	1.14*
	(0.76)	(2.51)	(2.51)
/			
var(M1[school])	0.32	0.34	0.34
	(1.61)	(1.62)	(1.62)
var(e.Intrinsic)	0.97***	0.97***	0.97***
	(14.95)	(14.95)	(14.95)
var(e.Extrinsic)	0.93***	0.93***	0.93***
	(14.95)	(14.95)	(14.95)
var(e.Costs)	0.94***	0.94***	0.94***
	(14.95)	(14.95)	(14.95)
var(e.Belonging)	0.92***	0.92***	0.92***
	(14.95)	(14.95)	(14.95)
var(e.Ability S. C.)	0.99***	0.99***	0.99***
	(14.95)	(14.95)	(14.95)
var(e.Abilitygap)	0.74***	0.77***	0.75***
	(14.95)	(14.95)	(14.95)
var(e.Hum)	2.60***	2.61***	2.61***

	(12.01)	(12.01)	(12.01)
var(e.Ecomath)	2.02***	2.03***	2.03***
	(12.05)	(12.05)	(12.05)
var(e.Stem)	1.95***	1.96***	1.96***
	(12.16)	(12.16)	(12.16)
N	**463**	**463**	**463**
AIC	**11833.58**	**11843.58**	**11830.21**
BIC	**12069.43**	**12050.47**	**12041.23**
Lrtest vs No random effect	LR chi2(1) = 79.20 (Prob > chi2 = 0.0000	LR chi2(1) = 85.43 Prob > chi2 = 0.0000	LR chi2(1) = 85.43 Prob > chi2 = 0.0000

Notes:
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4 presents the results of the models illustrated in the previous section: model 1 in the first column, model 2 in the second, and model 3 in the last, respectively. As explained in the methodological session, all indicators were standardized to enable the comparison of coefficients.

Panel A shows the impact of stereotypes on choice determinants according to gender, for each model. Since the first level is the same for all models: the values of the coefficients (and the respective comment) do not change across models. The coefficients for intrinsic value are not significant for males, indicating that stereotypes do not affect their enjoyment of mathematics or economics. However, for females, stereotypes have a significant negative effect on their liking of math. Regarding extrinsic value, males with stereotypes have a higher perceived utility linked to economics, while the opposite is observed for females. Stereotypes do not affect costs for either gender. Comparing the magnitude across indicators, stereotypes have the strongest impact on extrinsic value (utility) for both females and males, and on intrinsic value for females.

In this panel, it is noteworthy that, except for cost, the dummy ‘Female’ is never significant. This contrasts with the results of the t-test reported in Table 2 and highlights that being female, once we control for other determinants, does not inherently lead to lower subjective values. Rather, the significance of the gender-stereotype interaction suggests that stereotypes impact on female students choice determinants, rather than gender per se.

Panel B reports the effects of subjective indicators on the ability gap, for all models. Model 1 shows that intrinsic motivation and subjective perception of one's abilities are the only indicators that affect the performance gap, with similar magnitudes. Model 2 distinguishes between STV and ES variables, and proposes that the Ability Gap is exclusively influenced by the ability self-concept resulting in a strong and significative incidence of the ability self-concept on the ability gap. Model 3 combines the theoretical and empirical justifications of the first two models and shows that both intrinsic motivation and ability self-concept significantly affect the ability gap, with a higher total magnitude.

Finally, Panel C examines the impact of various determinants (intrinsic and extrinsic motivations, costs, belonging, and ability self-concept) as well as performance gaps, on enrolment intentions in humanistic, economics and mathematics, and STEM paths.

Our data shows that neither the performance gap nor most of the determinants significantly impact on the choice of humanistic paths. This lack of significance may be explained by the fact that subjective values, which are tailored for the mathematical and quantitative domains, may be less effective at explaining humanities path choices as through our essay. However, we do find that extrinsic motivation negatively affects the selection of a humanistic path, indicating that individuals who feel that quantitative and economic pathways are important for their personal and professional future, are less likely to pursue humanities degrees.

For Ecomath choices, a higher sense of belonging and extrinsic motivation increase the likelihood of enrolment, while intrinsic motivation has a significative effect only when the subjective task value is separated from the expectation of success and the direct effect of Ability self-concept on choices intentions is removed. For STEM choices, a strong appreciation and enjoyment of mathematics are the key factor.

The likelihood-ratio test confirms the necessity of a school random intercept for all models. Model 3, which includes all determinants, has the smallest AIC and BIC values, indicating that it is the best model. The coefficients are consistent across models and exhibit little variation.

In conclusion, our findings suggest that the presence of gender stereotypes negatively impacts most of the subjective values associated with the mathematical and economic domains. Indicators of intrinsic motivation and ability self-concept influence the gap between economic-mathematic and humanistic skills, while the sense of belonging and extrinsic and intrinsic motivations enhance enrolment intention towards majors in economics, maths, and statistics.

In the next chapter, devoted to conclusions, we reflect on the results obtained, comparing them with the literature and analysing which among them are in line with the literature and which constitute the main innovations.

7. CONCLUSIONS

Our findings shed important light, not only on the presence of stereotypes but also on the mechanism that convert stereotypes in biased choices.

Our study's findings are in line with prior research that utilized the IAT, including Carlana & Corno (2021), Carlana (2019), García-Montalvon (2020), and Nosek et al. (2009), which also identified the presence of implicit stereotypes associating gender to some academic disciplines. Specifically, the positive average score of the IAT test (0.40 for females and 0.14 for males) indicates that students regardless of their gender tend to associate females with humanities and males with math and economics. Therefore, we reject the null hypothesis for **H_1**, which confirms that students have embedded implicit stereotypes. Our findings also confirm **H_2**, as the absolute values of stereotypes and the significance of the t-test show that gender is affected differently by stereotypes. In our study, the female association with humanities is almost twice as high as the male association with math and economics.

A remarkable result arises from Panel A in table 4 that appears to be in contrast with the descriptive statistics reported in Table 2 and underscores that being female is not inherently linked to lower subjective values. Instead, our results suggest that it is the stereotypes that females are burdened with that influence choice determinants, rather than gender itself. So, accepting or rejecting **H_3** depends on its interpretation: If it is indeed true that female students, on average, have statistically lower subjective values, it is important to note that this outcome is not intrinsically linked to gender. Rather, it is perpetuated by the stereotypes. In other words, these stereotypes serve as a vehicle for the negative impact on subjective values, rather than gender itself being the sole determinant. Therefore, based on the more complex analysis, we reject this hypothesis as there is no direct impact of the female variable. This results have important implication in directly dealing with stereotypes to address the problem of lack of females in STEM and quantitative carrers.

Hypothesis **H_4** is supported by both the bivariate correlation and the results of panel A of GSEM. Specifically, the results show that as females' stereotype increases, their levels of extrinsic and intrinsic motivation decrease, as well as their sense of belonging. In contrast, male students exhibit an inverse bivariate correlation and positive coefficients in GSEM, significant only for extrinsic motivation, sense of belonging, and ability self-concept. As expected the results of males are in contrast to those of females: a high level of stereotyping in them causes positive feelings about the field of study.

As regards, **H_5**, not all subjective indicators impact performance, but rather intrinsic motivation and ability self-perception have a strong positive effect of roughly the same magnitude. It is important to note that, in our model, performance is not viewed as an absolute value, but rather as a skill gap between humanities and quantitative subjects. While the link between ability self-concept and personal achievement may seem straightforward and obvious, it is worth noting that intrinsic motivation is also a variable that displays strong evidence of its influence on school performance (Vansteenkiste et al. 2004; Orlicki 2019; Donze & Gunnes 2011; Deci, Cascio and Krussel 1975; Vallerand and Bissonnette 1992).

Our study supports the null hypothesis for **H_6** rejecting the fact that performance affect choices. Despite the pairwise correlation revealing a significant relationship between the Ability Gap and choices, this influence becomes insignificant in the GSEM model. This discrepancy may be since Ability Gap in the pairwise correlation also keeps the effect of subjective values (as they are related). Upon separating the various impacts in the GSEM model, we observe that students' subjective beliefs can have a greater impact on their actual choices than their actual skills. Finally, $\mathbf{H_7}$ is confirmed for certain determinants, such as the sense of belonging and intrinsic and extrinsic motivation.

To summarize the main takeaway from this analysis: we confirm that women do not inherently have values that steer them away from quantitative fields of study. It is not a biological matter of gender but rather it is the stereotype they are burdened with that influences the determinants of their choice and consequently their performance. Additionally, our analysis confirms Eccles' model theory that subjective determinants are highly impactful on choices and performances rather than the observed ability by itself.

Interestingly, cost indicators, such as fear, anxiety, and nervousness, appear to affect female students independently of stereotypes. These indicators are also the only ones that do not impact the ability gap or academic choices. However, it is unclear whether this effect is due to biological factors such as hormonal differences and genetic predispositions in females or to other societal constructs or beliefs that are not captured by our model, such as other forms of stereotypes beyond the association between gender and field of study. Furthermore, we lack sufficient information to determine whether this feeling is specific to economics and mathematics subjects or if it applies to all academic subjects, as we only explore costs for economics and mathematics. Despite these uncertainties, it is a commonly observed finding in research that females report higher levels of anxiety or nervousness than males when performing academic tasks especially if quantitative or math related (Rogowska 2020; Andrew 2012; Trifoni 2011; Cassady et al. 2002, Misra et al. 2000).

Based on the findings presented, it is recommended that policy interventions be developed and implemented to reduce stereotypes and/or their impact on academic choices and performance. This can be achieved by educating students, teachers, and parents about the presence of implicit stereotypes and useful actions aimed at addressing them. For example, the use of female role models has been found to be effective in lowering stereotypes (De Gioannis et all; 2023; Porter and Serra 2020; Breda et al. 2020; Carrell et. al. 2020; Redmond and Gutke 2019; Jethwani, et al. 2017). Schools can also encourage and support female students to pursue STEM subjects by providing them with resources and opportunities, such as scholarships, internships, and summer programs Additionally, dedicating some enrolment spots to highly quantitative courses and having a good plan to attract female students can help address the imbalances caused by stereotypes. Leveraging intrinsic and extrinsic motivation as well as the sense of belonging can also help achieve greater balance in choices.

Additionally, addressing the negative impact of anxiety and nervousness through counselling services and mindfulness practices can be beneficial but even more important is to better investigate which are the origins of this feeling. Further research is also necessary to better understand the mechanisms behind the persistence of stereotypes and the impact of anxiety and nervousness on academic performance and choice, which can inform future policy interventions and improve academic outcomes for all students.

There are limitations to our research, including that we only studied schools in a Northern district of Italy and collected enrolment intentions, which do not necessarily reflect actual enrolment. Additionally, our model assumes that stereotypes are exogenous, but endogeneity may be a concern, and future research could explore this further using instrumental variables such as gender awareness and guidance activities carried out within the classes or the institution. Information on these activities can be found on the *Scuole In Chiaro* portal: https://cercalatuascuola.istruzione.it/cercalatuascuola/.

Further development of this research could involve exploring the severity of the stereotype experienced by females: our analysis shows that females suffer from stereotypes in a much more severe form, which is almost 3 times greater than that experienced by males. It would be interesting to investigate this dimension further, starting with verifying whether it is unique to our sample. Secondly, although we have included the school intercept and therefore control for the school in which the students are enrolled, we know that students make their first important choice when they choose their high school. It would be interesting to replicate the experiment for that age group or, even better, to consider a panel analysis that follows the students over time. And finally, it could be interesting to explore if the same model holds for the humanistic dimension when the subjective indicators used are tailored to the humanistic domain.

8. LITERATURE

Addabbo, T. Strozzi, C. Tasselli, C. (2022) “Gender inequalities in college major choices. The role of explicit and implicit stereotypes” Working paper, DEMB WORKING PAPER SERIES, Dipartimento di Economia Marco Biagi – Università degli Studi di Modena e Reggio Emilia. https://doi.org/10.25431/11380_1284504.

Adelman, H. S. (1978). The concept of intrinsic motivation: Implications for practice and research with the learning disabled. Learning Disability Quarterly, 1, 43-54.

Akaike, H. (1987). Factor analysis and AIC. Psychometrika 52: 317–332. https://doi.org/10.1007/BF02294359.

Allan W., Eccles, J S. (2020) 35 years of research on students' subjective task values and motivation: A look back and a look forward, Advances in Motivation Science, Elsevier, Volume 7, Pages 161-198, ISBN 9780128196342, https://doi.org/10.1016/bs.adms.2019.05.002.

AlmaLaurea. (2022). XXIV Indagine sul Profilo dei Laureati 2021. AlmaLaurea. https://www.almalaurea.it/universita/indagini/laureati/profile.

AlmaLaurea. (2022b). XXIV Condizione occupazionale dei Laureati 2021. AlmaLaurea. https://www.almalaurea.it/universita/indagini/laureati/profile.

Arcidiacono, P., & Nicholson, S. (2005). Peer effects in medical school. Journal of Labor Economics, 23(3), 457-483.

Augustyniak, Robert A., Adrienne Z. Ables, Philip Guilford, Heidi L. Lujan, Ronald N. Cortright, and Stephen E. DiCarlo. 2016. ‘Intrinsic Motivation: An Overlooked Component for Student Success’. *Advances in Physiology Education* 40 (4): 465–66. https://doi.org/10.1152/advan.00072.2016.

Bandura A. (1997). Self-Efficacy: The Exercise of Control. New York: Freeman.

Bandura, A., Barbaranelli C., Caprara G.V., and Pastorelli C. (2001). ‘Self-Efficacy Beliefs as Shapers of Children’s Aspirations and Career Trajectories’. Child Development 72 (1): 187–206. https://doi.org/10.1111/1467-8624.00273.De Gioannis, E. (2022a). ‘Implicit Gender-Science

Stereotypes and College-Major Intentions of Italian Adolescents'. *Social Psychology of Education*, July. https://doi.org/10.1007/s11218-022-09709-3.

Bertocchi, G., Luca B., and Marina M. (2022). 'Adams and Eves: The Gender Gap in Economics Majors'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4114415.

Carlana, M. , and Corno L. (2021). 'Parents and Peers: Gender Stereotypes in the Field of Study'. 16582. *CEPR Discussion Papers*. CEPR Discussion Papers. C.E.P.R. Discussion Papers. https://ideas.repec.org/p/cpr/ceprdp/16582.html.

Carlana, M. (2019). 'Implicit Stereotypes: Evidence from Teachers' Gender Bias'. *The Quarterly Journal of Economics* 134 (3): 1163–1224. https://doi.org/10.1093/qje/qjz008.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. Contemporary educational psychology, 27(2), 270-295.

Carnevale, A. P., Smith, N., & Strohl, J. (2013). The economic value of college majors. Georgetown University Center on Education and the Workforce.

Compton, Ying Li, Sok-Hyon Kang, and Zinan Zhu. (2019). 'Gender Stereotyping by Location, Female Director Appointments and Financial Performance'. Journal of Business Ethics 160 (2): 445–62. https://doi.org/10.1007/s10551-018-3942-y.

Corno, L. , La Ferrara E., Burns, J.. (2019). 'Interaction, Stereotypes and Performance. Evidence from South Africa'. W19/03. *IFS Working Papers*. IFS Working Papers. Institute for Fiscal Studies. https://ideas.repec.org/p/ifs/ifsewp/19-03.html.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests psychometrika, 16(3), 297-334.

Deci, E. L., Cascio, W. F., & Krusell, J. (1975). Cognitive evaluation theory and some comments on the Calder and Staw critique.

De Gioannis, E. (2022b). 'The Conundrum of Gender-Science Stereotypes: A Review and Discussion of Measurements'. *Quality & Quantity*, August. https://doi.org/10.1007/s11135-022-01512-8.

De Gioannis, E., Pasin G.L. and Squazzoni F. (2023). 'Empowering Women in STEM: A Scoping Review of Interventions with Role Models'. *International Journal of Science Education, Part B*, January, 1–15. https://doi.org/10.1080/21548455.2022.2162832.

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2018). I like them, but do they like me? Ethnic and gender differences in adolescents' perceived popularity with peers and career aspirations. Developmental Psychology, 54(6), 1166-1181. https://doi.org/10.1037/dev0000513

Doepke, M., Sorrenti, G., Zilibotti, F. (2019). The Economics of Parenting, *Annual Review of Economics*, 11 (1): 55–84

Donze J. & Gunnes T., (2011). "Should economists listen to educational psychologists? Some economics of student motivation," *Discussion Papers 662*, Statistics Norway, Research Department .

Dynan, K. E., & Rouse, C. E. (1997). The underrepresentation of women in economics: A study of undergraduate economics students. The Journal of Economic Education, 28(4), 350-368.

Eagly, A. H., Wood, W., & Diekman, A. B. (2019). Social role theory of sex differences and similarities: A current appraisal. In Advances in experimental social psychology (Vol. 61, pp. 1-74). Academic Press.

Eccles, J. S. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. International Journal of Behavioral Development, 35(3), 195-2

Eccles, J., and Wigfield A. (2002). 'Motivational Beliefs, Values, and Goals'. *Annual Review of Psychology* 53 (1): 109–32. https://doi.org/10.1146/annurev.psych.53.100901.135153.

Eccles, J. S. (1994). Understanding women's educational and occupational choices: Applying the Eccles et al. Model of Achievement-Related Choices. Psychology of Women Quarterly, 18(4), 585–609. https://doi.org/10.1111/j.1471-6402.1994.tb01049.x.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. Psychology of Women Quarterly, 11(2), 135–172. https://doi.org/10.1111/j.1471-6402.1987.tb00781.x .

Eccles, J. (1983). 'Expectancies, Values, and Academic Behaviors'. In *J. T. Spence (Ed.*), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: J. H. Freeman

Feather NT. (1992). 'Values, valences, expectations, and actions.' J. Soc. Issues (48):109–24

Feather NT. (1988). 'Values, valences, and course 93nrolment: testing the role of personal values within an expectancy-value framework.' J. Educ. Psychol. (80):381–91Fong, Carlton, and Kristen Kremer. (2019). 'An Expectancy-Value Approach to Math Underachievement:

Examining High School Achievement, College Attendance, and STEM Interest'. *Gifted Child Quarterly* 64 (November). https://doi.org/10.1177/0016986219890599.

Froiland, J. M., & Worrell, F. C. (2016). Intrinsic motivation, learning goals, engagement, and achievement in a diverse high school. Psychology in the Schools, 53(3), 321–336. https://doi.org/10.1002/pits.21901

Heyman, G.D., Dweck, C.S. Achievement goals and intrinsic motivation: Their relation and their role in adaptive motivation. Motiv Emot 16, 231–247 (1992). https://doi.org/10.1007/BF00991653.

Hood, Michelle, Peter A. Creed, and David L. Neumann. (2012). 'Using the Expectancy Value Model of Motivation to Understand the Relationship between Student Attitudes and Achievement in Statistics'. *Statistics Education Research Journal* 11: 72–85. https://doi.org/10.52041/serj.v11i2.330.

Holcomb, Tim R., James G. Combs, David G. Sirmon, and Jennifer Sexton. (2010). 'Modeling Levels and Time in Entrepreneurship Research: An Illustration with Growth Strategies and Post-IPO Performance'. Organizational Research Methods 13 (2): 348–89.

García-Montalvon J & Alimonti D. & Reiland S. & Vernos I., (2020). "Gender Stereotype and the Scientific Career of Women: Evidence from Biomedical Research Centers," Working Papers 1212, Barcelona School of Economics.

Giustinelli, P. (2016). 'Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of the High School Track', *International Economic Review*, 57 (2): 573–602, https://doi.org/10.1111/iere.12168.

Good, C., Rattan, A., & Dweck, C. S. (2012). 'Why do women opt out? Sense of belonging and women's representation in mathematics'. Journal of Personality and Social Psychology, 102(4), 700–717. https://doi.org/10.1037/a0026659 .

Granato, S. (2018). Gender inequalities and scarring effects in school to work transitions.

Greene, W. H. (2018). Econometric Analysis. 8th ed. New York: Pearson, ISBN-13: 978-0-13-446136-6.

Greenwald, A. G., Nosek B. A., and Banaji M. R. (2003). ‘Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm.’ Journal of Personality and Social Psychology 85 (2): 197–216. https://doi.org/10.1037/0022-3514.85.2.197.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). ‘Measuring individual differences in implicit cognition: The implicit association test’. Journal of Personality and Social Psychology, 74(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464.

Kirkham J. & Elaine Chapman, E. (2020). "Gendered Decision-Making About Mathematics Courses: Contributions of Self-Perceptions, Domain-Perceptions, and Sociocultural Factors," *Journal of Educational and Developmental Psychology*, Canadian Center of Science and Education, vol. 10(1), pages 1-43, May.

Lane, Kristin A., Jin X. Goh, and Erin Driver-Linn. (2012). ‘Implicit Science Stereotypes Mediate the Relationship between Gender and Academic Participation’. Sex Roles 66 (3–4): 220–34. https://doi.org/10.1007/s11199-011-0036-z.

Lizzeri, A., Siniscalchi, M. (2008). ‘Parental Guidance and Supervised Learning’, *The Quarterly Journal of Economics*, 123 (3): 1161–1195. https://doi.org/10.1162/qjec.2008.123.3.1161

Misra, R., & McKean, M. (2000). College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. American journal of Health studies, 16(1), 41.

Nosek, Brian A., Frederick L. Smyth, N. Sriram, Nicole M. Lindner, Thierry Devos, Alfonso Ayala, Yoav Bar-Anan, et al. (2009). ‘National Differences in Gender–Science Stereotypes Predict National Sex Differences in Science and Math Achievement.’ PNAS Proceedings of the National Academy of Sciences of the United States of America 106: 10593–97. https://doi.org/10.1073/pnas.0809921106.

Nunnally, J. C. (1978). Psychometric theory: New York : McGraw-Hill, c1978. 2d ed

Orlicki, E. (2019). "Formation of Human Capital in Knowledge Intensive Service Activities," *Revista de Economía Política de Buenos Aires, Universidad de Buenos Aires. Facultad de Ciencias Económicas*., issue 19, pages 101-136, December

Perez, Tony, Jennifer G. Cromley, and Avi Kaplan. (2014). 'The Role of Identity Development, Values, and Costs in College STEM Retention.' *Journal of Educational Psychology* 106: 315–29. https://doi.org/10.1037/a0034027.

Raftery, A. E. (1993). Bayesian model selection in structural equation models. Reprinted in Testing Structural Equation Models, ed. K. A. Bollen and J. S. Long, pp. 163–180. Newbury Park, CA: SAGE.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary educational psychology, 25(1), 54-67.

Rogowska, A. M., Kuśnierz, C., & Bokszczanin, A. (2020). Examining anxiety, life satisfaction, general health, stress and coping styles during COVID-19 pandemic in Polish sample of university students. Psychology Research and Behavior Management, 797-811.

Rokeach, M. (1979). 'From the individual to institutional values with special reference to the values of science.' In M. Rokeach (Ed.), Understanding human values (pp. 47-70). New York, NY: Free Press.

Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2009). Girls' and boys' reading and writing motivation: Their independent contributions toward academic achievement. Sex Roles, 60(1-2), 23-35. https://doi.org/10.1007/s11199-008-9471-1.

Shyr, Wen-Jye, Ying-Ming Hsieh, and Ching-Huei Chen. (2021). 'The Effects of Peer-Based Instant Response System to Promote Learning Performance, Intrinsic Motivation and Self-Efficacy'. Sustainability 13 (8): 1–13.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. 'Stereotype Threat and Women's Math Performance'. *Journal of Experimental Social Psychology* 35 (1): 4–28. https://doi.org/10.1006/jesp.1998.1373.

Stinebrickner, T. (2014). A major in science? Initial beliefs and final outcomes for college major and dropout. National Bureau of Economic Research.

StataCorp.(2021). 'Stata structural equation modelling reference manual' *Stata: Release 17. Statistical Software*. College Station, TX: StataCorp LLC. ISBN-13: 978-1-59718-347-5.

Steele, Claude M. (1997). 'A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance.' *American Psychologist* 52: 613–29. https://doi.org/10.1037/0003-066X.52.6.613.

Stoevenbelt, A. H., Paulette C. Flore, I. Schwabe, and Jelte M. Wicherts. 2022. 'The Uniformity of Stereotype Threat: Analyzing the Moderating Effects of Premeasured Performance'. Intelligence 93 (July): 101655. https://doi.org/10.1016/j.intell.2022.101655.

Tobin J (1958). 'Estimation of relationship for limited dependent variables.' Econometrica, 26, 24–36.

Tomasetto C., Mirisola A., Galdi S., and Cadinu, M. (2015). 'Parents' Math–Gender Stereotypes, Children's Self-Perception of Ability, and Children's Appraisal of Parents' Evaluations in 6-Year-Olds'. *Contemporary Educational Psychology* 42: 186–98. https://doi.org/10.1016/j.cedpsych.2015.06.007.

Trifoni, A., & Shahini, M. (2011). How does exam anxiety affect the performance of university students. Mediterranean journal of social sciences, 2(2), 93-100.

Vallerand, R. J., & Blssonnette, R. (1992). Intrinsic, extrinsic, and a motivational styles as predictors of behavior: A prospective study. Journal of personality, 60(3), 599-620.

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. Journal of personality and social psychology, 87(2), 246-260.

9. APPENDIX

Appendix A1 – dependent variables elaboration

The classification of the 14 university options into macro groups is based on the disciplinary area groupings of the Italian Ministry of Education, University and Research (MIUR). The only exception at Miur grouping concerns our main dependent variable of interest "EcoMath" which also combines economics with two majors derived from hard stems (mathematics and statistics).

Table A1 – aggregation of choice options

Macro-field	**subgroups**
HUM	political science, sociology (futurechoice_13) law (futurechoice_14) I literature, philosophy, languages, pedagogy and psychology (futurechoice_15)
HEALTH AND MEDICINE	Agricultural or veterinary sciences (futurechoice_5) medicine and dentistry (futurechoice_6) nursing and physiotherapy (futurechoice_7)
STEM (Hard)	Physics or chemistry (futurechoice_3) Pursue studies at a faculty of biology science or pharmacy (futurechoice_4) Engineering (futurechoice_8) Study at a university of architecture and urban planning (futurechoice_9)
ECOMATH	Mathematics (futurechoice_2) Economics and marketing (futurechoice_10) Economics and finance (futurechoice_11) Statistics (futurechoice_12)

ESSAY N° 3

Team-Based Learning as an innovative teaching methodology: assessing gender inclusiveness in a quantitative economic subject.

Chiara Tasselli[a] , Tindara Addabbo[b], Filippo Damiani[c]

a University of Modena and Reggio Emilia, Marco Biagi Foundation
Email: chiara.tasselli@unimore.it

b University of Modena and Reggio Emilia, Gender Equality Observatory, Marco Biagi Foundation
Address: Viale Berengario 51, 41121, Modena, Italy
Email: tindara.addabbo@unimore.it

c University of Modena and Reggio Emilia, Modena, Italy
Email: filippo.damiani@unimore.it

ABSTRACT

This paper analyses the impact on students' performance of the introduction of team-based learning (TBL, for short) teaching methodology. TBL has proven to be a powerful and versatile teaching strategy that enables teachers to take small group learnings to a new level of effectiveness according to the empirical evidence of its implementation worldwide since 2020.

The analysis involves various cohorts of macroeconomics students at the University of Modena and Reggio Emilia Marco Biagi Department of Economics (DEMB). We exploit the structural break in the academic year 2017/2018 in which TBL was introduced in the Macroeconomics Course at the second year of the Bachelor Programme in Business and management to test the impact of TBL teaching methods in learning outcomes also considering students' heterogeneity.

Within this perspective of inclusiveness, special attention is given to gender bias in TBL effectiveness and outcomes in Macroeconomics measured by using Macroeconomics exam grades.

Comparisons between groups were made using econometric and statistical analysis on a rich and self-constructed database. The latter, composed using multiple administrative and primary data sources, allows controlling for student socio-demographic characteristics and their academic careers. The econometric analysis starts with a *multivariate regression* to estimate the TBL's effect on grades in macroeconomics, then moves to a *probit analysis* that returns the probability of passing the exam and concludes with a *Cragg model* (two-part Hurdle model) which offers improved estimates by correcting for the censored dependent variable. Regardless of gender, a positive impact of TBL on students' performance is detected. Treated, especially if female, improves macroeconomic scores, and this is consistent with the literature reviewed. On the other hand, males benefit more from the treatment through a significant increase in the likelihood of passing the exam.

1. INTRODUCTION

Previous studies in the education field have shown that gender plays a decisive role in students' academic enrolment and performance in economics, scientific and technological universities. Even if empirical evidence (Castagnetti & Rosti 2009) and data[36] show that women outperform men in Italian Universities, female low self-confidence in quantitative and scientific courses is found to affect their access to employment and horizontal segregation in the labour market. That structural problem -in addition, to perpetuating gender inequality in this field and making difficult to reach SDG 5 results in a massive loss of talent in the economic system.

This paper focuses on the impact of a change in teaching methodology, from traditional teacher-centred lecturing to a methodology that has been proven to promote active learning and teamwork, namely Team Based Learning (TBL). TBL has been developed by Michaelsen in the late 1970s, and has been increasingly used in the US since the 1980s in a variety of disciplines in tertiary education through its application to economics has been limited (Cagliesi & Ghanei, 2022). TBL has been introduced in an undergraduate course in Macroeconomics within a wider project carried out by a public university in the North of Italy based on its expected positive impact on the students' soft skills in problem-solving and teamwork development.

As assessed by Simkins, Maier, & Ruder (2021), TBL intentionally promotes learning strategies that learning sciences research identified as highly effective to create powerful learning environments for students. The attention paid to the group's composition resulting in within groups diversity also on the ground of gender allowed us to test also its impact on inclusion.

The analysis compares students' performance through a consistent and robust estimator for models with censored data using Cragg's model *(or two-part, Hurdle model)* on a sample of 711 students[37] attending a macroeconomics course at the Marco Biagi Department of Economics of the University of Modena & Reggio Emilia.

It can be, therefore hypothesized that:

H_1 = *attending Team-Based Learning Lessons produces better learning outcomes.*

H_2 = *students react differently to treatment depending on their gender.*

---

https://www2.almalaurea.it/cgi-php/universita/statistiche/tendine.php?LANG=it&config=profilo.

[37] For a total amount of 1,024 exam attempts.

H_3 = *Female performance in Macroeconomics is lower.*

H_4 = *TBL could help overcome gender differences in macroeconomics.*

We do try to solve these research questions or at least try to get a clearer view of the relationship between students, their progression in the learning process and macroeconomics classes (also from a gender perspective).

The paper is structured as follows: Section 2 is dedicated to the literature review and presents the principles of Team-Based Learning that we adopted it in our courses. Section 3 presents the sample selection and distribution (3.1) together whit a detailed and in-depth insight into data and variables description (3.2) which continues in the annexes section. Then, the methodology followed and the results of the estimated models are presented in section 4. Section 5 provides some concluding remarks.

2. TEAM BASED LEARNING METHODOLOGY AND EXPECTED OUTCOMES

The focus of this paper is on the impact of a particular methodology that has been recognised in the literature as able to develop students' active engagement, specific soft-skills and, in its implementation, allows a high degree of inclusiveness: Team based learning (TBL).

TBL has been developed by Michaelsen in the late 1970s, and has been increasingly used in the US since the 1980s in a variety of disciplines in tertiary education though its application to economics has been limited (Cagliesi & Ghanei, 2022).

Michaelsen *et al.* (2004a), described TBL as an unusually powerful and versatile teaching strategy that enables teachers to take small group learnings to a new level of effectiveness. TBL group work has been found to be powerful in improving the ability of the students to apply course contents since, during TBL activities, the development of self-managed learning teams is promoted (Michaelsen & Sweet, 2008; Michaelsen, Davidson & Major, 2014).

TBL teams composition and in their duration play a crucial role in the efficacy of the approach and its inclusive content. TBL teams are formed and the membership of the groups must be kept stable during the whole term to allow team development (Michaelsen, Watson, & Sharp, 1991). Care must be taken on the composition of the groups since it has indeed been demonstrated that the most effective results are obtained in groups with the most diverse composition possible (Parmelee & Michaelsen, 2010; Phillips et al., 2008), which means that groups are deliberately formed to be diverse and cohesive (Kathleen & Odell 2018). The dimension of groups is of 5-7 members in order to ensure the group dimension that is considered efficient to face the variety of decision-based tasks encountered during TBL implementation (Michaelsen et al. 2004b).

TBL can be considered as a student-centred class methodology. Students are assigned course materials before a teaching session (flipped classroom) to be able to apply in classes their self-gained knowledge (Balan et al. 2015). In-class activities are typically based on the Readiness Assurance Process (RAP), which consists of two Readiness Assurance Tests (RAT) in which the students should answer the same questions first individually (iRAT, Step one), and then as a team (tRAT, Step two).
Then, after the instructor's clarification lecture on the first set of questions, students work again on a team application (tAPP, Step 3). As stated by Espey (2018):

“Significant problems engage students in concrete examples so they understand the usefulness of the course concepts. Specific choices require teams to take a position, sometimes also requiring them to support that position with a short rationale of their choice. Forcing all students to confront the same problem enables them to better engage with each other across teams, while simultaneous reporting precludes teams from simply agreeing with the majority of others, forcing them to decide before knowing what other groups will say.” [Espey, 2018, p.10]

The fourth part of the activities consists of peer assessment and feedback, leading to students' evaluation of their teammates (Step 4); this last part is fundamental to enhance the ability to work together and positively contribute to the team (Michaelsen, Davidson & Major, 2014) and to avoid freeriding (Hettler, 2015).

While frequently implemented in a face-to-face classroom, TBL has received limited attention in the online learning environment were geographically distributed, and asynchronous learning poses challenges to its fundamental design (Goh et al., 2020). Virtual reality could be a platform to provide the engaging elements of TBL, without students needing to be physically present in the same room. It has the potential to be a useful tool for online, distance TBL (Coyne et al., 2018).

Amongst the positive impact of TBL, the literature has shown increased students' engagement both in class and out of class (Imazeki 2015; Espey, 2012; Ruder, Maier & Simkins 2021) and increased attendance (Abio et al. 2019). Evidence has been provided on a positive impact of the adoption of TBL in the percentage of show up of students at the final exam and in their rate of success in passing the exam for students re-taking a subject (Abio et al. 2019) and for students in STEMM courses (Parappilly et al., 2021). Evaluation of the TBL implementation in principles of microeconomics and quantitative methods courses as compared to lecture-based instruction, allowed Hettler (2015) to detect differences in the outcome of TBL on the exam scores for different groups of students namely, the minority and first-generation college students status show a positive and significant marginal impact on exam score in TBL sections thus supporting the hypothesis that TBL can have a higher impact on groups that are typically disadvantaged. Cagliesi & Ghanei (2022) found evidence of a positive impact of TBL on grades in economic courses and a reduction in the attainment gap for Black, Asian, and minority ethnic students.

In terms of the efficacy of TBL methodology to foster inclusion, not only evidence has been provided of the reduction in the achievement gaps for minorities attending courses using TBL sessions, but also evidence has been provided on the TBL approach to be more attractive for female and non-white students (Clerici-Arias, 2021).

Another line of investigation on TBL evaluation concerns the impact of teams' characteristics on teams or individual outcomes or behaviour in teams on individual outcomes. Espey (2018) analyses what measurable characteristics of teams influence team and individual performance on the comprehensive final exam. The latter has been found to be positively affected both for men and for women by a more equal gender distribution within TBL groups. Espey (2022) shows evidence of a positive impact on final exam scores of increased effort or engagement in team-based activities.

3. TEAM BASED LEARNING IMPLEMENTATION AND THE GENERATED DATA

TBL methodologies have already been adopted in the University of Modena and Reggio Emilia (Unimore) in 2017 within the project "Didactics for competencies" involving about 2,000 students in the experimentation showing a positive impact on the development of soft skills considered fundamental in business contexts (De Santis *et al*, 2019, Bellini et al, 2020). Through contacts with stakeholders (companies, public and private bodies, the tertiary sector) Unimore identified the two soft skills that at the beginning of the project were the most demanded in the labour market: problem-solving, i.e. an approach to work that, by identifying priorities and critical issues, allows the identification of the best possible solutions to problems; teamwork, i.e. the willingness to work and collaborate with others, having the desire to build positive relationships aimed at achieving the assigned task. TBL has then been chosen as a methodology able to develop these soft skills and implemented in the academic year 2017/2018 in 16 courses with 16 control courses that have allowed an evaluation of the impact of TBL on students' soft skills. Instructors and tutors involved in the TBL courses have been involved in a training course to acquire knowledge on TBL methodology and on how to restructure their syllabus. A community of practices has then been built within Unimore in strict collaboration with the Italian National TBL Community of practices.

The undergraduate course in Macroeconomics analysed in this paper, has been involved from the very beginning of the TBL implementation and the data collected refer in total to 891 students (1,345 including those who repeated the exam) who attended the course from the academic year 2016/2017 (when TBL has not yet been implemented) to the academic year 2020/2021.

To ensure diversity, the groups were created using G(roup)Rumbler, an algorithm developed by Prof. Malcolm K. Sparrow in 2011 to maximize the mixing of students across the class (Sparrow, 2011). The variables that have been used in this implementation of the GRumbler to form the TBL groups have been collected throughout a survey run before TBL classes in each academic year and refer to gender, age, origin, type of high school attended, grades in Math and Microeconomics, students' attitude in team working, personal characteristics, etc. The goal was to allow within-group diversity in line with what has been found to increase the effectiveness of TBL in developing teamwork and problem-solving and also to have a positive impact on inclusiveness. The group membership has been kept permanent with semester-long teams.

The implementation of TBL in the Macroeconomics semester course is structured in 30 lectures using active learning techniques and six TBL units with partial pre-class assignments following the Readiness Assurance Process four steps structure described in Section 2.

3.1 Sample.

To avoid any possible contamination of the data with the occurrence of the pandemic, we decided to cut the sample in February 2020. Before this date, both the teaching and examination methods remained practically unchanged, except for the introduction of the TBL in the academic year 2017/2018.[38] The final sample, therefore, consists of 711 students and 1,024 exams attempts.

Students are cohorts attending the second year of the Undergraduate Course in Macroeconomics from the academic year 2016/2017 till 2020/2021[39]. The lectures take place in the first Semester of the academic year from September to December. A total of 6 exams run each academic year: 2 in the Winter Session, 3 in the Summer Session and 1 in the Fall Session.

Table 1 shows the composition of the final sample also including the exam session undertaken. The table shows that before the introduction of TBL, students preferred to sit for the second winter call, whereas from 2018 onwards they show up predominantly to the first call of the Winter Session.

Out of the 1,024 examination tests, 439 were carried by female students and the remaining 585 by male students. The distribution between sessions tends to be concentrated in the winter session the closest to the semester when the course is taught. Female students have an average attendance rate[40] in TBL nearly 5 percentage points higher than that of men.

A deeper insight concerning the distribution of the sample, its relation to the treatment and its respective characteristics will be addressed later.

[38] Moreover, for the last exam sessions, it would not be possible to use the whole list of covariates proposed in the analysis as the databases are currently being updated.

[39] Only until the February session because the Summer sessions suffer from pandemic interference.

[40] Percentage gap estimated only on students who attended classes from the introduction of TBL and, hence, got the opportunity to participate.

Tab 1 – Sample by gender and date of exam

	MALE		FEMALE		
SESSION	*Treat=0*	*Treat=1*	*Treat=0*	*Treat=1*	TOTAL
Jan 17	33 100%	X	36 100%	X	69
Feb 17	76 100%	X	46 100%	X	122
July 17	16 100%	X	11 100%	X	27
Sept 17	11 100%	X	12 100%	X	23
Gen 18	19 28.8 %	47 71.2%	9 17.3%	43 82.7%	118
Feb 18	21 36.8%	36 63.2%	9 18.4%	40 81.6%	106
May 18	3 75%	1 25%	6 50%	6 50%	16
June 18	7 63.6%	4 36.4%	4 40%	6 60%	21
July 18	8 57.1%	6 42.9%	6 40%	9 60%	29
Jan 19	16 21.1%	60 78.9%	7 18%	32 82.1%	115
Feb 19	21 38.2%	34 61.8%	9 30%	21 70%	85
May 19	2 15.4%	11 84.6%	2 25%	6 75%	21
June 19	6 60%	4 40%	5 55.6%	4 44.4%	19
July 19	10 62.5%	6 37.5%	3 60%	2 40%	21
Sept 19	9 56.3%	7 43.8%	5 50%	5 50%	26
Jan 20	17 23%	57 77.0%	21 35.6%	38 64.4%	133
Feb 20	16 43.2%	21 56.8%	13 36.1%	23 64%	73
Total	291	294	204	235	1024

Source: self-elaboration on primary & administrative data.

3.2 Data and variables description:

To investigate the gender differences in macroeconomics among undergraduate students analysed, multiple data sources have been merged.

Administrative data have been downloaded for the purpose from the Unimore student management system[41], results of intermediate tests collected by the professor and socio-demographic and behavioural covariates obtained by submitting students a questionnaire.

Measuring Variables

Dependent and independent variables are defined as follows (a more detailed description of the variables is provided in *Table A1* in Appendix 1):

The dependent variables used to represent students' academic performance are two: a continuous variable which reports the students' final grade in macroeconomics (*Mark)* and a dummy variable (*Pass*) stating if the student passed or failed the exam.

It is important to stress that the selection of the *Mark* variable implies our sample being classified as censored from the below (and above) sample. The latter is representative of the population because all students who have Attempted the macroeconomics exam at least once are in the sample, but the mean of the dependent variable is not because we cannot observe students' marks if they fail the exam as we do not know their true performance if they succeeded. This means that the variable has a lower bound set on the score of 17, and for students who cannot reach it we cannot observe the actual performance. The same applies to the highest extremity of the distribution where there is an upper bound at 30 cum laude (we cannot observe the real mark over 30).

On the other hand, several **explanatory variables** are used in this study.

Some related to students' academic paths like:

- *i)* The university entrance score at TOLC[42] in Math *(MathAbility)*
- *ii)* The university entrance score at TOLC in Logic *(LogicAbility)*
- *iii)* The university entrance score at TOLC in reading comprehension *(ComprehensionAbility)*

  Points (i, ii a and iii) are considered proxies of ability before entering university.
- *iv)* if they have already Attempted the test *(Retaker)*
- *v)* whether they attend TBL classes or not *(TBL)* and the *v)* TBL dosage *(Dosage)*.
- *vi)* The number of credits obtained in the first year *(Credits)*
- *vii)* outside prescribed time students *(OverTimeGrad)*

[41] *Student Management System* (Sistema per la gestione studenti: ESSE3) is one of the "core" services of Cineca's suite of products to support "Didactics and Students" in the university environment. First of all ESSE3 allows to manage (and follow) the entire academic' *"life cycle"* of the student.

[42] TEST ONLINE CISIA: more detail in tab. 1 in addendum.

viii) the average of all the exams taken by the students during their academic career subdivided into 3 macro-groups[43](whose disaggregation is detailed in *Table A2* in section 7 – addendum):

a. Highlyquantitative

b. Slightlyquantitative

c. Nonquantitative

Other covariates relating to students' sociodemographic characteristics are also included, namely:

ix) gender *(Female)*

x) *(LowIncome)* as a low family income could adversely affect school performance

xi) *(Native)* Italian nationality

In addition, time-fixed effects account for all unobservable factors that are changing across sessions. This method is useful for increasing the adjusted R-squared because it allows each session to have a customised coefficient that increases the goodness of fit.

Table A1 in the Appendix provides an in-depth and detailed description of the variables grouped by:

- Dependent variables
- Independent variables
- Minor variables used for descriptive statistics or to provide an in-depth view of the sample

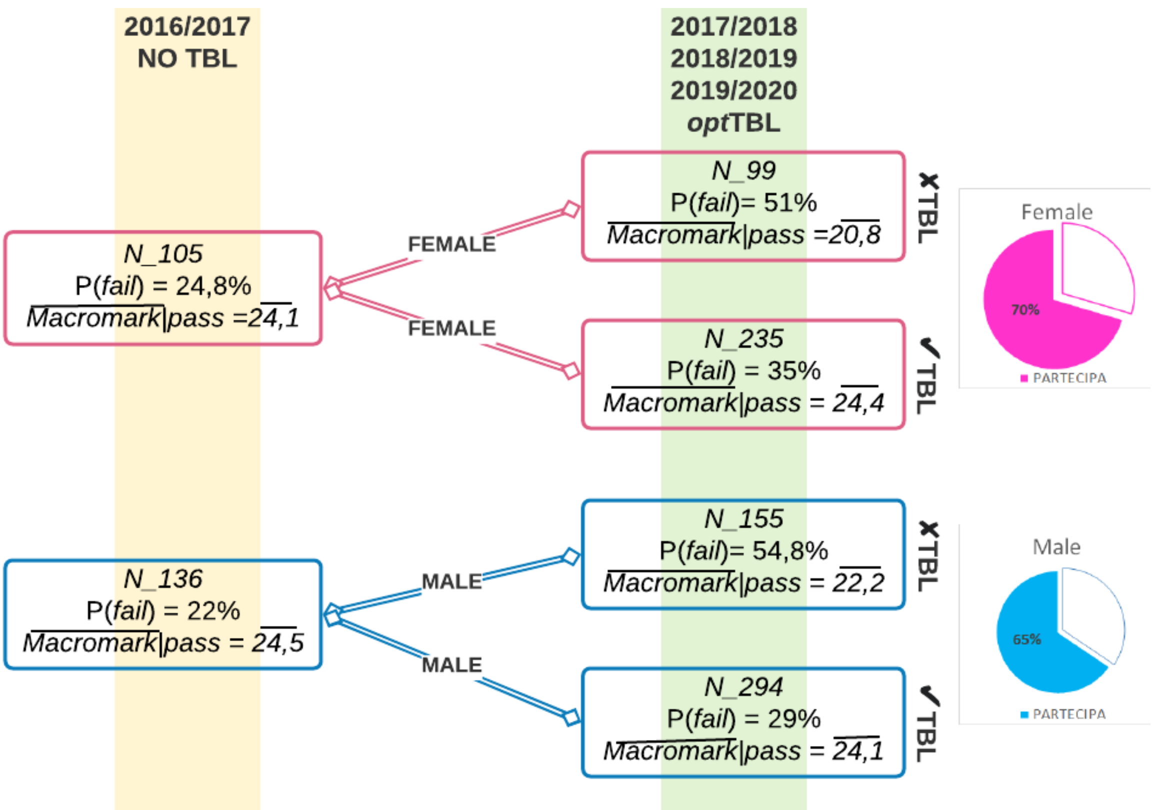
[43] The applied procedure considers that each student may have experienced a different academic pathway for the selection of optional exams. The proposed aggregation system succeeds in preserving all the shades of the paths without generating an excessive loss of observations (see section 3.2 for more detail).

4 EMPIRICAL ANALYSIS

4.1 Research design and implementation

The research design reflects the methodologies adopted in the Introductory macroeconomics course that is the object of this study. Before the academic year, 2017-2018 the course was held mainly in a traditional lecture-based format and, thereafter the same instructor changed the structure of the course by adopting the TBL approach. Groups of 5-6 members have been formed according to the GRumbler algorithm referred to in Section 3 of this paper, which considered socio-demographic characteristics such as gender, ethnicity, openness, and scholastic skills. Team membership has been kept stable throughout the duration of the semester and students worked together to solve the T-Rat and the case study (T-App) meanwhile they face the I-Rat and the teammates' evaluation individually. For each year in which TBL was implemented, the intervention dosage consists of 6 sessions -lasting an hour and a half each – distributed in the semester, the rest consisted of lectures and classes where active participation of students was required as in the instructor's style of lecturing. We, therefore, use the structural break from lecture-based to TBL-based course of the same course and the same instructor to evaluate TBL impact on students' achievements controlling for a set of variables that includes also the students' cohort.

Figure 1 – details on Treatment and ITT.



Source: self-elaboration on primary data-

Tools lucid Chart [pie charts are excel imported]

Figure1 shows the distribution of the sample by gender year of the exam and education path covered. What stands out in the figure is that the raw average mark in macroeconomics (calculated on positive marks only) for females who had not experienced TBL is lower than the male counterpart. Vice versa females who attend TBL succeed in getting a higher mark compared to males. This finding is consistent with the literature surveyed in Section 2, showing a reduction of certain groups of students' gap in achievements when the TBL methodology is adopted.

Figure 1 also reveals that female students show a higher participation rate in TBL than male students again a result consistent with the literature and in line with the inclusive scope of the methodology application.

4.2 Methodology

The statistical analysis was carried out using the econometric package STATA/BE 17.0 in which:

i) a *regression model* has been estimated to detect what were the most significant variables for our analysis and their best combination both for the goodness of fit and the coefficient strength (definition of the model).

ii) a *probit model* has been implemented to find out the probability of passing the macroeconomics exam depending on the characteristics of the students and the academic pathway taken.

iii) a *Cragg's model* (two-part, Hurdle model) has been estimated to obtain the best possible fit for this sample in which the dependent variable is censored from both the bottom and the top. This model is a modified version of the Tobit model (Tobin 1958) and is preferred to the latter following a likelihoods log test.

All these steps will be developed in detail in paragraph 4.4 dedicated to the econometric analysis.[44]

The methodological approach was organised as follows: initially, intensive data merging and data cleaning work were carried out: several databases of primary data (own and third-party collections) were obtained[45] and merged with administrative data. To summarise briefly data collection also involved the Introductory macroeconomics instructor, current and past Introductory Macroeconomics tutors, the teaching coordinator responsible for the administrative data sets.

[44] STATA/BE 17.0 has been used to carry out the estimation.

[45] Variables used to generate *IratScore, Dosage* (and *Treat*) originate from six different year databases of primary data as a result of downloads from the Moodle platform. The process repeats each year that TBL was implemented for a total 18 *subdatabase* for the current sample (2021 if we consider the initial sample, which also included the 2020/2021 academic year).

The set of variables (including TBL experience) whose impact on students' achievements has been shown in the literature has been taken into account and included in the estimated models considering their distribution, the presence of systematic missing for certain groups[46], over time homogeneity[47] and whether their inclusion would impose an excessive reduction of the sample together with their significance in explaining the gender gap in the outcomes.[48] Some variables were excluded from the final model because they do not have significant coefficients and do not contribute to increasing the adjusted-R^2 (*NUTS1 dummies*, *EnrollGap*, etc..). On the other hand, variables such as *NearbyHighSchool*, *MathAbility*, *LogicAbility* and *LowIncome* were retained even though their coefficients were not significant as they contribute to increase the fit of the model.

The selection criterion for the final model was based on a comparison of the following requirements: ① having as greater adjusted R^2 as possible, ② significance of the coefficients (β) and ③ limitation of missing values. If some of them conflict, the adjusted R^2 has been privileged, if the submodels conflict the fit of the female-student model estimate was favoured.

4.3 descriptive statistic and first findings

Table 2 shows descriptive statistics for the covariates characterizing our sample. Panel A is dedicated to continuous variables, meanwhile, panel B is to dichotomous ones. Both panels consist of two subsections that make gender comparisons between the control (1) and the treated (2) group The typical self-selecting student as treated has a higher average in all subject types (high, middle and Non quantitative); has a short time gap from graduation to university enrolment, and earned more credits in the first year of the undergraduate course attended.

Table 2 – Dependent and Independent Variable Descriptive Statistics

Panel A.1: Continuous Variable for controls

	(1) MALE		(2) FEMALE		(3) T-TEST $(\overline{x_m} - \overline{x_f})$	
	mean	sd	mean	sd	b	t
Mark	20.97	4.62	20.70	4.42	0.28	(0.67)
Highlyquantitative	23.36	2.86	23.50	3.10	-0.13	(-0.47)
Slightlyquantitative	22.59	2.94	22.44	2.72	0.15	(0.51)
Nonquantitative	23.60	2.43	23.91	2.45	-0.31	(-1.34)
Tolc[a]	16.45	5.18	12.39	5.16	4.06***	(8.46)
Comprehensionability	5.85	1.88	4.74	2.09	1.11***	(5.27)

[46] *IraTot* (missing in Treat== 0), *left* and *OverTimeGrad* not reliable for more recent years.

[47] *TOLC; EnglishAbility;etc…*

[48] See Table 2 in section 4.3 and Table A1 in the Appendix for more details.

Mathability	4.52	3.02	3.04	2.51	1.48***	(5.18)
Logicability	6.25	2.24	5.40	2.52	0.85***	(3.36)
Tmaxingl[B]	20.84	4.06	21.60	4.19	-0.76	(-0.85)
Dosage	0.05	0.46	0.15	0.63	-0.10	(-1.84)
Iratscore[C]	0.12	1.12	0.80	3.41	-0.68**	(-2.74)
Attempts	1.52	0.97	1.49	0.84	0.03	(0.37)
Gapfromdiploma	0.38	2.52	0.50	1.60	-0.12	(-0.66)
Credits	38.11	16.69	37.75	15.80	0.36	(0.24)
N	291		204		495	

Panel A.2: Continuous Variable for treated

	(1)		(2)		(3)	
	MALE		FEMALE		T-TEST $(\overline{x_m} - \overline{x_f})$	
	mean	sd	mean	sd	b	t
Mark	22.02	4.90	21.80	4.99	0.22	(0.50)
highlyquantitative	24.73	3.04	24.77	3.25	-0.03	(-0.11)
slightlyquantitative	24.21	2.97	23.56	3.07	0.65*	(2.00)
nonquantitative	24.70	2.51	24.60	2.38	0.10	(0.44)
TOLC[A]	17.19	6.10	14.24	5.53	2.95***	(5.73)
ComprehensionAbility	5.56	2.40	5.06	2.22	0.50*	(2.26)
MathAbility	5.02	3.01	3.57	3.06	1.44***	(4.93)
LogicAbility	6.85	2.53	5.35	2.46	1.49***	(6.23)
Tmaxingl[B]	20.75	5.10	20.27	5.21	0.49	(0.96)
Dosage	5.82	0.39	5.80	0.40	0.02	(0.69)
IratScore	16.08	8.37	16.75	7.61	-0.67	(-0.96)
Attempts	1.34	0.63	1.44	0.73	-0.11	(-1.75)
GapfromDiploma	0.21	0.73	0.24	0.79	-0.03	(-0.45)
Credits	39.91	16.06	35.62	15.99	4.29**	(3.02)
N	294		235		529	

Panel B.1.: Dichotomous Variable for control

	(1)		(2)		(3)	
	MALE		FEMALE		T-TEST$(\overline{x_m} - \overline{x_f})$	
	mean	sd	mean	sd	b	t
Pass	0.60	0.49	0.63	0.48	-0.02	(-0.51)
Retaker	0.32	0.47	0.32	0.47	-0.00	(-0.09)
cred40	0.51	0.50	0.46	0.50	0.05	(1.15)
DropOut	0.04	0.20	0.06	0.25	-0.02	(-1.03)
native	0.96	0.20	0.85	0.36	0.11***	(3.80)
LowIcome	0.06	0.23	0.25	0.43	-0.19***	(-5.74)
MiddleIcome	0.13	0.34	0.10	0.31	0.03	(0.93)
HighIcome	0.81	0.39	0.65	0.48	0.16***	(4.00)
OverTimeGrad	0.38	0.49	0.42	0.49	-0.04	(-0.90)
Winter	0.75	0.43	0.74	0.44	0.02	(0.43)
Northeast	0.87	0.34	0.87	0.34	0.00	(0.06)
Northwest	0.01	0.12	0.02	0.14	-0.01	(-0.49)
Center	0.04	0.20	0.08	0.28	-0.04	(-1.86)
South&Islands	0.06	0.23	0.07	0.26	-0.02	(-0.66)
NearbyHighSchool	0.88	0.32	0.86	0.35	0.02	(0.68)

N	291		204		495	

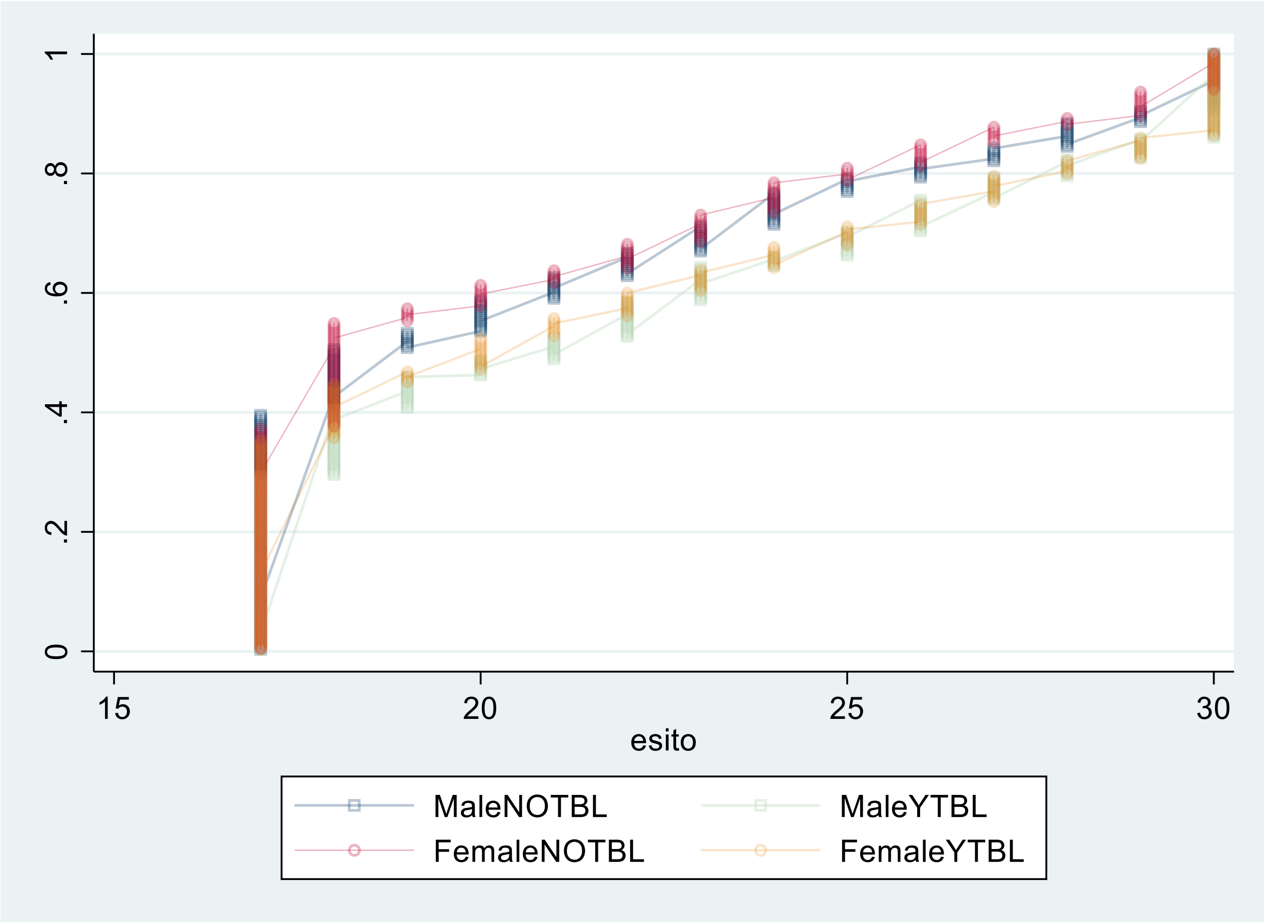
Panel B.1.: Dichotomous Variable for treated

	(1) MALE		(2) FEMALE		(3) T-TEST $(\overline{x_m} - \overline{x_f})$	
	mean	sd	mean	sd	b	t
Pass	0.71	0.46	0.65	0.48	0.06	(1.48)
Retaker	0.26	0.44	0.33	0.47	-0.07	(-1.73)
cred40	0.56	0.50	0.44	0.50	0.12^{**}	(2.83)
DropOut	0.02	0.13	0.01	0.11	0.00	(0.40)
native	0.92	0.28	0.89	0.32	0.03	(1.12)
LowIcome	0.18	0.38	0.24	0.43	-0.06	(-1.72)
MiddleIcome	0.07	0.25	0.06	0.25	0.00	(0.04)
HighIcome	0.75	0.43	0.69	0.46	0.06	(1.56)
OverTimeGrad	0.06	0.24	0.07	0.25	-0.01	(-0.48)
Winter	0.87	0.34	0.84	0.37	0.03	(0.93)
Northeast	0.84	0.37	0.87	0.34	-0.03	(-1.01)
Northwest	0.02	0.13	0.02	0.14	-0.00	(-0.35)
Center	0.06	0.25	0.06	0.24	0.01	(0.24)
South&Islands	0.10	0.29	0.07	0.26	0.02	(0.95)
NearbyHighSchool	0.84	0.36	0.88	0.32	-0.04	(-1.29)
N	294		235		529	

Source: self-elaboration on primary & administrative data.
Notes: t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[A] TOLC = has changed its composition for students enrolled from 2017 onwards (the English evaluation was introduced).
[B] variable not homogeneous in the sample (missing not at random) it is detectable only for students enrolled from 2017 onwards
[C] I-rat score could exist even if Treat is zero. It belongs to students who participated at TBL without reaching the minimum treatment dosage.

This condition is verified for both male and female students and could suggest that students who self-select into treatment are the most deeply motivated. Regardless of the cluster the majority of students have attended a high school in the same region as the university (neighbourhood proxy) and have a stable household financial situation. In the last column of Table 2 (panel A and Panel B) a t-test (Two-sample t-test with equal variances) on covariates was also included to see if the covariates assume significant gender differences within *Treat* groups.

Figure 2 – Cumulative function of the macroeconomic outcome by gender and TBL attendance.



Source: self-elaboration on primary data
Tool STATA/BE 17.0

T-test does not reveal any particular gender differences in the covariates except for TOLC scores (*ComprehensionAbility, MathAbility, LogicAbility)* . Descriptive statistics show that, although female students score significantly worse on the entrance test, they manage to achieve an academic performance in the same line or even higher than that of male students. The worse female students' performance in the entrance test could also be due to the type of question framing. Already previous literature emphasizes how the multiple-choice question type results in disadvantages for females (Reardon et al., 2018) and how the higher risk perceived on average by female students in answering this type of questions impacts on the lower their lower performance (Baldiga, 2014; Karimi & Biria, 2017).

Finally, Figure 2 presents the cumulative distribution of macroeconomics grades of the macroeconomic grades of 4 clusters just presented in table 2.

What stands out in the table is the truncation that occurs in both the lower and upper bounds: grades have a higher concentration at the edges of the distribution, as all students who score below 18 or above 30 enter the dataset as 17 or 30 respectively. This is why a Hurdle model can be preferred in the estimation to the regression model that does not consider the double truncation in the grades. Another interesting aspect is the left shift of the cumulative function relating to clusters of students who did not experience the TBL. The more the cumulative function is shifted to the left, the more individuals in the group are concentrated in low scores. The figure shows that in the 4 clusters:

- keeping gender constant, having participated in TBL leads to a rightward shift of the curve (better performance)
- keeping the treatment (TBL participation) constant, the male cumulative curve is always to the right of the female one.

4.4 Econometric models

i) Regression model

The estimated model (1) whose results are shown in Table 4 has been obtained after more than 90 trials of models' estimation.

(1)

$$Mark_i = \beta_0 + \beta_1 Treat + \beta_2 Female + D_3 Session + \beta_4 Credits + \beta_5 Reteker + \beta_6 Native + \beta_7 LowIncome + \beta_8 Highlyquantitative + \beta_9 Slightlyquantitative + \beta_{10} nonquantitative + \beta_{11} ComprehensionAbility + \beta_{12} MathAbility + \beta_{13} LogicAbility + \varepsilon_i$$

(1a)

$$Mark_i \,|_{female} = \beta_0 + \beta_1 Treat + D_2 Session + \beta_3 Credits + \beta_4 Retaker + \beta_5 Native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8 Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11} MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

(1b)

$$Mark_i \,|_{male} = \beta_0 + \beta_1 TBLcutoff + D_2 session + \beta_3 credY1 + \beta_4 rep + \beta_5 native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8 Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11} MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

$Mark_i$ is the Macroeconomic performance obtained by a female/male student, 30 is the maximum grade a student can achieve and 18 is the minimum grade that a student can get.

β_1 is the coefficient related to the TBL effect on macroeconomics grades, D_2 regards a set of 16 dichotomous time variables that have been included and account for all unobservable factors that are changing across sessions.

β_7, β_8 and β_9 are coefficients indicating the effects of the student's average[49] exam in the highly quantitative, slightly quantitative and non-quantitative disciplines respectively.

β_{10}, β_{11} and β_{12} return the effect of entry capabilities (before university) i.e. the results of the Entry test in reading comprehension (ComprehensionAbility) in math (MathAbility) and in logic (LogicAbility) on *Mark* and finally ε_i contains all errors resulting from omitted variables and the respective loss of information.

Table 4 illustrates the STATA outcome for models **1, 1a** and **1b** before commenting them, it is important to remember that the regression regarded a dependent variable in which all insufficient marks (which are unobserved) were set to 17.

The regression in Table 4 shows that participation in TBL is highly significant for all students and exams marks are 1.7 points higher than for those who do not participate. The main benefits are for female-students (in line with the literature surveyed in Section 2), who get almost 3 more points in the exam by taking part in TBL. Being born in Italy seems to be important for male-students but not relevant for female-students' mark in macroeconomics. The positive and significant impact of the *Highlyquantitative, Slightlyquantitative and Nonquantitative* variables coefficients are expected as the level of preparation and ability of the students is linked to all past exams' performance however the impact is not higher for courses with a high quantitative content. The positive and significant coefficient of *Retaker*, a variable that takes the value of one when the exam has been retaken, can be found only for female students, this result can be linked to a strategy that is more frequent for female students to sit for the exam as an Attempts to acquire familiarity with the exam structure and then accept only the highest mark and would require higher investigation (also interacting the variable with the TBL experience to test the positive impact of TBL on re-takers achievements detected in the literature surveyed in Section 2).

[49] As (due to student's self-selection) each student may have experienced a different academic pathway, subgroups means were computed *ignoring missing* values; for example, if three exams are specified and, some students, select only two, in those observations *newvar* will contain the mean of the two variables that do exist. This procedure makes it possible to minimise the generation of missing values that would otherwise be caused by different academic choices. At the same time it allows a large number of examination marks to be considered. Mean is computed on the mark that have been recorded in the students' academic record reflecting only marks over 17 that have not been rejected by the students after the exam (University of Modena and Reggio Emilia does not include a rule of acceptance of the mark obtained in the exam and it is possible for students to reject the mark obtained and re-take the exam, a maximum of 4 trials of the exams are allowed in an academic year), insufficient marks as well as rejected marks by the students are not detectable.

Table 4 – Results of the estimation of models 1,1a and 1b.

	(1) ALL	(2) FEMALE	(3) MALE
Treat	1.689^{***}	2.809^{***}	1.341^{*}
	(3.52)	(3.77)	(2.11)
female	-0.0286		
	(-0.08)		
Credits	0.0509^{***}	0.0676^{***}	0.0323
	(3.91)	(3.90)	(1.65)
native	1.241	-0.701	3.388^{**}
	(1.95)	(-0.94)	(3.17)
LowIncome	-0.173	0.426	-0.0604
	(-0.35)	(0.69)	(-0.08)
highlyquantitative	0.251^{***}	0.181	0.274^{**}
	(3.32)	(1.65)	(2.63)
slightlyquantitative	0.363^{***}	0.536^{***}	0.332^{**}
	(4.80)	(4.79)	(3.13)
nonquantitative	0.449^{***}	0.428^{**}	0.474^{**}
	(4.18)	(2.77)	(3.17)
Retaker	0.814	1.677^{**}	0.497
	(1.94)	(2.92)	(0.82)
NearbyHighSchool	-0.703	0.379	-1.512
	(-1.27)	(0.50)	(-1.86)
ComprehensionAbility	-0.0113	0.262^{*}	-0.228^{*}
	(-0.14)	(2.18)	(-1.99)
MathAbility	0.0525	0.0717	0.0398
	(0.81)	(0.77)	(0.44)
LogicAbility	-0.0160	0.0379	-0.0462
	(-0.21)	(0.38)	(-0.39)
Session FE	YES	YES	YES
Constant	-4.677^{*}	-8.096^{*}	-4.410
	(-2.11)	(-2.50)	(-1.42)
N	585	243	342

Source: self-elaboration on primary & administrative data.
Notes: *t* statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

ii) Probit model

(1) *Binary outcome of the dependent variable* $Pass \begin{cases} 0 \text{ if student fail} \\ 1 \text{ if student pass} \end{cases}$

(2)

$$P(pass = 1|X) = \beta_0 + \beta_1 Treat + \beta_2 Female + D_3 Session + \beta_4 Credits + \beta_5 Reteker + \beta_6 Native + \beta_7 LowIncome + \beta_8 Highlyquantitative + \beta_9\, Slightlyquantitative + \beta_{10} nonquantitative + \beta_{11} ComprehensionAbility + \beta_{12}\, MathAbility + \beta_{13} LogicAbility + \varepsilon_i$$

(2a)

$$P(pass = 1|X_{female}) = \beta_0 + \beta_1 Treat + D_2 Session + \beta_3 Credits + \beta_4 Retaker + \beta_5 Native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8\, Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11}\, MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

(2b)

$$P(pass = 1|X_{male}) = \beta_0 + \beta_1 Treat + D_2 Session + \beta_3 Credits + \beta_4 Retaker + \beta_5 Native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8 Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11} MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

The outcome of probit estimations ***2***, ***2a*** and ***2b*** are displayed in table 5 where marginal effects computed at the means of the variables are displayed. Once again, the diversity of the results by gender can be observed. TBL treatment seems to have a positive and significant impact on male-students' pass probability. In addition, it is interesting to highlight that having participated to the TBL seems to be the only variable determining the probability of males passing. This means that participating to the TBL becomes more important than the males' abilities (Highlyquantitative, Slightlyquantitative, Nonquantitative) and their diligence (number of credits acquired in the first year).

Table 5 –Marginal effects at means of model 2, 2a and 2b.

	(1) ALL	(2) FEMALE	(3) MALE
Treat	0.286	-0.260	0.642**
	(1.67)	(-0.82)	(2.83)
female	0.144		
	(0.99)		
Credits	0.0177***	0.0250**	0.0142*
	(3.70)	(3.15)	(2.01)
native	0.180	-0.339	0.705
	(0.81)	(-1.01)	(1.92)
LowIncome	-0.198	-0.112	-0.172
	(-1.10)	(-0.41)	(-0.59)
highlyquantitative	0.0157	-0.001	0.0307
	(0.57)	(-0.02)	(0.83)
slightlyquantitative	0.0638*	0.134*	0.0594
	(2.21)	(2.47)	(1.49)
nonquantitative	0.123**	0.208**	0.0657
	(2.92)	(2.67)	(1.15)
Retaker	0.223	0.766**	0.006
	(1.45)	(2.83)	(0.03)
NearbyHighSchool	-0.216	-0.225	-0.473
	(-0.90)	(-0.58)	(-1.16)
ComprehensionAbility	0.00303	0.141*	-0.0659
	(0.09)	(2.29)	(-1.48)
MathAbility	0.0377	0.0189	0.0480
	(1.45)	(0.39)	(1.38)
LogicAbility	0.0532	0.0380	0.0825
	(1.79)	(0.78)	(1.85)
SESSION FE	YES	YES	YES
_cons	-4.802***	-8.074***	-3.668**
	(-5.08)	(-4.41)	(-2.87)
N	585	243	334
CORRECTLY CLASSIFIED	77,44%	81,48%	78,74%

Source: self-elaboration on primary & administrative data.
Note: Statistical significance at the 1%, 5% and 10% levels is denoted by ***, **, *
Note: dy/dx for factor levels is the discrete change from the base level.

iii) Cragg's model (two-part, hurdle model)

As anticipated in section 4.2 in this analysis we face a censored sample since the value of the dependent variable is not detectable below (or above) a defined threshold. conveniently, unlike the truncated sample, the censored sample is representative of the population[50] because all observations are included, only the dependent variable suffers losses of information.

[50] Macroeconomics students in our case.

In our analysis, the test scores of the students who pass are detectable and range from 18 to 30, while we cannot observe the scores of the students who fail the exam (they could have scored 2 points as 17 and for us, it is only known that they do not get a sufficient evaluation). The model has, therefore, a lower limit at 17 and -to increase its accuracy- we decided to also consider an upper limit at 30 as it is not possible to distinguish the different marks within excellences.

Tobin (1958) develops a Tobit model that provides consistent and efficient estimators under this restrictive assumption on the dependent variable.

Eq. 4, following Tobin's specification, shows the dependent variable's processing within the model. The actual value for $MacroMark$ is observed if the latent variable $MacroMark^*$ is between 18 and 30 meanwhile lower limit is observed for the censored from below observations and the upper limit is observed for the censored from above observations.

$$\textbf{\textit{Eq 4}} \quad Mark_i = \begin{cases} Mark_i^* & if \quad 18 \leq Mark_i^* < 30 \\ 17 & if \quad Mark_i^* < 18 \\ 30 & if \quad Mark_i^* \geq 30 \end{cases}$$

(3)

$$Mark_i = \beta_0 + \beta_1 Treat + \beta_2 Female + D_3 Session + \beta_4 Credits + \beta_5 Reteker + \beta_6 Native + \beta_7 LowIncome + \beta_8 Highlyquantitative + \beta_9\, Slightlyquantitative + \beta_{10} nonquantitative + \beta_{11} ComprehensionAbility + \beta_{12}\, MathAbility + \beta_{13} LogicAbility + \varepsilon_i$$

(3a)

$$Mark_i\,|_{female} = \beta_0 + \beta_1 Treat + D_2 Session + \beta_3 Credits + \beta_4 Retaker + \beta_5 Native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8\, Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11}\, MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

(3b)

$$Mark_i\,|_{male} = \beta_0 + \beta_1 TBLcutoff + D_2 session + \beta_3 credY1 + \beta_4 rep + \beta_5 native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8\, Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11}\, MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

The combination of covariates used to estimate this functional form is the same as in ***reg***. ***1***, ***1a*** and ***1b*** with the distinction that the dependent variable (*Mark*) is not a simple continuous variable but assume the latent ($Mark$) form in ***Eq 4*** to obtain model ***3***, ***3a*** and ***3b.***

The Tobit model was applied in two-step (rather than one) relaxing the assumption that the discrete event and the continuous event are the same, allowing different coefficients for the 1 probability of passing the exam[51] and for the ② continuous grade variable once a passing grade has been achieved *(Cragg's model* - a Tobit variant*).*

Table 6 – Output of models 3, 3a and 3b.

Coefficients are omitted and marginals effects are displayed:
dydx(*) at means predict(e(17,.)), E(Mark| Mark>17), predict(e(17,.))

	(1) ALL	(2) FEMALE	(3) MALE
Treat	4.889**	7.624***	2.265
	(3.09)	(4.02)	(1.50)
Female	0.124		
	(0.14)		
Credits	0.101**	0.070*	0.059
	(2.66)	(2.07)	(1.42)
Native	2.384	-0.639	7.470*
	(1.46)	(-0.56)	(2.40)
LowIncome	0.359	1.650	0.040
	(0.29)	(1.52)	(0.02)
Highlyquantitative	0.355	0.404	0.276
	(1.72)	(1.86)	(1.15)
Slightlyquantitative	0.434*	0.468*	0.446
	(2.13)	(2.18)	(1.85)
Nonquantitative	0.952**	0.565*	1.016**
	(3.19)	(2.10)	(2.99)
Retaker	1.049	0.625	1.656
	(0.93)	(0.55)	(1.25)
NearbyHighSchool	0.349	2.045	-1.199
	(0.26)	(1.51)	(-0.75)
ComprehensionAbility	0.002	0.175	-0.269
	(0.01)	(0.86)	(-1.19)
MathAbility	-0.054	0.055	-0.088
	(-0.35)	(0.38)	(-0.50)
LogicAbility	-0.232	-0.026	-0.385
	(-1.20)	(-0.15)	(-1.58)
SESSION FE	YES	YES	YES
constant	-25.086*	-15.459*	-25.882*
	(-2.51)	(-2.12)	(-2.31)
sigma	4.577***	3.201***	4.146***
	(8.80)	(8.91)	(7.84)
N	346	143	203

Source: self-elaboration on primary & administrative data.
Notes: *t* statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[51] Which we have already presented in eq. 2 and respective outputs in tab 5.

The decision to opt for the Cragg's model was taken following a dedicated test for the best fit as displayed in ***Eq 5***. We estimated separately tobit, probit, and truncated regression (Cragg's) models and derived their log-likelihoods to compute the following likelihood ratio statistic:

$$\textbf{\textit{Eq 5}} \quad \lambda = 2 * (\mathrm{LL}_{probit} + \mathrm{LL}_{truncreg} - \mathrm{LL}_{tobit})$$

Manual application of the formula is reported in addendum A5 in Appendix 1 where the chi-square test validates the best fit of the Cragg's model. This condition is verified both for the main model and for its gender disaggregation.

As Cragg's model develops in two consecutive stages, the output of table 6 must be analysed in light of the findings of the probability of being promoted in the Probit model (table 5). Table 6 displays the marginal effect computed *at means* of covariates rather than coefficients (β) because the latter indicates the estimate of the latent variable (*Mark*) whereas by giving an actual value (mean) to the covariate we can calculate the real $Mark$.

Observing the first column of table 6, we note that -once promoted- having participated in the TBL increases the examination' marks by a 3points, a result validated for each level of significance. Especially the female subgroup seems to weigh on the sample size and effect. The largest effect in the table is seen for females who were attended classes when TBL has been implemented: among the female students who did not fail, those who participated in TBL scored approximately 6 points higher than those who did not. This evidence is significant for each level. On the contrary for male-students, although positive, this variable is not significant. This result is in line with the literature results surveyed in Section 2 showing the higher impact of TBL on students' exam marks for certain groups of the students' population.

5. CONCLUSIONS AND FUTURE RESEARCH DEVELOPMENT

The primary aim of this essay is to evaluate the impact of Team Based Learning on students' achievements measured by the exam marks and the passing probability with reference to different cohorts of students attending the Introductory Macroeconomics Course in a Bachelor Degree Programme in a public University located in the North of Italy.

Multivariate analyses performed in this essay provide evidence on H1 hypothesis that attending Team-Based Learning Lessons produces better learning outcomes measured by the exam grade both by estimating an OLS model and even more by estimating a Hurdle Model consistently with what has been shown in the literature: Espey(2022) by analysing three different economics courses found that TBL produce better learning outcomes togheter with higher levels of engagement in team activities and also Cagliesi & Ghanei(2022) founds Their findings that TBL improved students' academic performance and reduced several achievement gaps in economics class

The models' estimation provides evidence of a different impact by gender to the introduction of TBL thus satisfying H2 hypothesis that students react differently to treatment depending on their gender. In fact we find a higher impact, in terms of grades, for female students, with no significant impact on their probability of passing the exam, while, on the contrary, male students' probability of passing the exam is strongly affected by TBL bearing a lower effect on grades.

The higher positive impact on female students' grade can be connected to the higher training that female students can benefit while attending TBL due to the structure of TBL tests more similar to the final exam structure. In fact the latter contains multiple choice questions characterized on average by a lower performance for female students, as shown by Griselda (2020); Karimi & Biria (2017), Baldiga, (2014)

Meanwhile, the two joint results (impact on grades and probability of passing) might suggest that the practice of TBL not only influences knowledge but can also affect students' behaviour.
The literature shows that female students, who are more risk-averse, tend to show up for the exam only if they are fully prepared, while male students, in contrast, tend more to "*try it*" (Dohmen, 2010; Croston 2009; Niederle 2007). Therefore we can claim that the TBL practice on both male and female students may have also acted on their approach to exams by empowering -or at least compensating- them.
With regards to the common belief that female students perform worse than male students in macroeconomics as a quantitative subject (H3 = Female performance in Macroeconomics is lower) the hypothesis is not confirmed by our results. In fact, although descriptive statistics (table 2, panel A) show that women score significantly lower on the entrance test (TOLC and any of its subgroups), they manage to achieve similar performance in line than their male counterparts (or even higher if they attended TBL). Once again, one possible explanation for this discrepancy is that the TOLC test is based on multiple-choice questions, in which female students on average tend to underperform. Supporting this hypothesis is the fact that female students also underperform in verbal comprehension in the entry test (TOLC), whereas they usually tend to be more talented in this latter. These results are remarkable as they suggest that TBL practice may help female students in overcoming their disadvantage in multiple-choice questions even in quantitative domains. This point is another research goal that we are trying to further qualifying by analysing the structural break of the Covid pandemic that caused the massive use of this testing modality. With regards to H4 hypothesis that TBL could help in overcoming gender differences in macroeconomics, **e**ven if

our analysis does not show significant gender differences in macroeconomics exam marks, participating in TBL, as tested by our estimation, has an extremely positive impact on female outcomes and this can help to prevent the gap from widening. In addition, as proved by the models estimated and referred to above, TBL teaching may have not only affected the economics grade but also female students' behaviours in approaching multiple choices tests. Observed changes in students' behaviour after TBL practices have also been observed by other authors, including Christensen *et al.* 2019; Lin 2019; Dearnley *et al.* 2018.

Collectively, these findings indicate a positive and meaningful correlation between TBL course attendance and performance in macroeconomics exams, while accounting for individual variables related to students' socio-demographic and cognitive skills. Additionally, these results reveal a shift in students' behaviour and exam strategies, which warrants further investigation through data collection on this dimension.

These findings have important implications for policy interventions aimed at improving macroeconomic education outcomes or -in general- mitigating gender imbalances in quantitative courses and/or multiple choice evaluation. Given the positive association between attending TBL's courses and exam performance, institutions are encouraged to promote the use of TBL in Macroeconomics courses. This may involve providing training to instructors on how to implement this teaching methodology effectively. Furthermore, universities could offer financial incentives or other forms of support to encourage instructors to adopt TBL, thereby promoting gender equality and improving learning outcomes for all students.

In order to reduce the penalties that multiple-choice tests have on female students, institutions should consider alternative assessment methods. Alternatively, they could provide training and resources to support female students in developing the skills and confidence needed to perform better on multiple-choice tests. The latter policy can also be pursued in high school in order to impact on female students' performance in multiple choice tests that usually characterize entrance tests for university courses with limited number of admitted students. Another option would be for professors to plan mixed exams that include both multiple-choice and other types of questions. These interventions could help reduce gender imbalances in assessment outcomes and support greater equity in academic performance. However, we are aware of the limitation of the present study. The TBL course attendance is not compulsory, and students can opt-out and not attend the TBL sessions or not attend the course but simply sit for the exam as not attending students. Most of the students who had the option of attending the TBL course did opt in however we could not exclude students' self-selection into treatment though it is conceivable to assume that the covariates included in the analyses help to correct for bias due to self-selection, but a stronger counterfactual group is needed. Further developments include the introduction of a parallel traditional course in Introductory Macroeconomics held by another Instructor by following a lecture-based approach without any opportunity to have TBL sessions, to improve the evaluation of the impact of TBL through the counterfactual or diff in diff techniques. Alternatively and/or additionally, it is suggested to include the Heckman (1979) correction for non-random selection in the treatment. Finally, the application of the Oaxaca decomposition (Blinder 1973; Oaxaca 1973) proposed by Bauer and Sinning (2010)[52,] can allow the detection of gender differential in Introductory Macroeconomics, not due to differences in the observed characteristics.

Further developments will include evaluation of other likely TBL outcomes as observed in the literature as the level of engagement in the learning process and development of problem-solving, and teamwork skills as well as more collaborative behaviour and degree of attractiveness of TBL courses for disadvantaged groups of the students' population. Moreover, viewing TBL also as an inclusive teaching methodology, we are planning to measure students' perceived sense of inclusion by including in our data collection validate scales regarding the "Sense of belonging" (Good et all. 2012) and Survey on Diversity, Equity and Inclusion.

Finally. further research could explore the potential effectiveness of TBL in other subject areas and courses, comparing it with other teaching techniques and their respective impact on students' learning outcomes.

[52] This method is preferred over the one implemented in Jann (2008) because Monte Carlo simulations demonstrate that in the case of censored dependent variables this decomposition method produces more reliable results than the conventional Blinder–Oaxaca decomposition for linear regression models (Bauer & Sinning, 2010).

6. REFERENCES

Abio, G., M. Alcaniz, M. Gomez-Puig, G. Rubert, M. Serrano, A. Stoyanova, and M. Vilalta-Bufi. (2019). Retaking a course in economics: Innovative teaching strategies to improve academic performance in groups of low-performing students. *Innovations in Education and Teaching International* 56 (2): 206–16.

Baldiga, K. (2014). Gender Differences in Willingness to Guess. Management Science 60 (2), 434-448. doi:10.1287/mnsc.2013.1776.

Ballester, Pilar, C. 'Factors That Influence Academic Performance: Analyzing Gender Differences in Accounting Students'. *Revista Educación* 43, no. 2 (2019): 29. https://doi.org/10.15517/revedu.v43i2.28916

Bauer, Thomas, & Mathias Sinning. «Blinder-Oaxaca decomposition for Tobit models». *Applied Economics* 42, n. 12 (2010): 1569–75. https://doi.org/10.1080/00036840701721612

Bellini, C.; De Santis, A.; K. Sannicandro, K.; Cecconi, L. (2020). Dalla formazione dei docenti alle competenze trasversali degli studenti: un progetto di faculty development presso Unimore, - in Faculty Development in Italia. Valorizzazione delle competenze didattiche dei docenti universitari, a cura di A. Lotti, P. A. Lampugnani.

Blinder, Alan S. 'Wage Discrimination: Reduced Form and Structural Estimates'. *Journal of Human Resources* 8, no. 4 (1973): 436–55.

Bilen, E., & Matros, A. (2021). Online cheating amid COVID-19. Journal of Economic Behavior & Organization 182, 196-211. doi:10.1016/j.jebo.2020.12.004.

Burke, William J. (2009). ‘Fitting and Interpreting Cragg’s Tobit Alternative Using Stata’. *The Stata Journal: Promoting Communications on Statistics and Stata* 9 (4): 584–92. https://doi.org/10.1177/1536867X0900900405.

Cagliesi, G. & Ghanei, M. (2022) ‘Team-Based Learning in Economics: Promoting Group Collaboration, Diversity and Inclusion’, *Journal of Economic Education*, 53(1), pp. 11–30.

Castagnetti, Carolina, and Luisa Rosti. (2009). ‘Effort Allocation in Tournaments: The Effect of Gender on Academic Performance in Italian Universities’. *Economics of Education Review* 28 (3): 357–69. https://doi.org/10.1016/j.econedurev.2008.06.004.

Clerici-Arias, M. (2020). Underrepresented groups in the economics major. The impact of using team-based learning in the principles course. Paper presented at the 2020 Annual Meeting of the Allied Social Science Associations, San Diego, CA, January 4.

Coyne, Leanne, Jody K. Takemoto, Brittany L. Parmentier, Thayer Merritt, and Rachel A. Sharpton. (2018). ‘Exploring Virtual Reality as a Platform for Distance Team-Based Learning’. *Currents in Pharmacy Teaching and Learning* 10 (10): 1384–90. https://doi.org/10.1016/j.cptl.2018.07.005.

Christensen, J., Harrison, J. L., Hollindale, J., & Wood, K. (2019). Implementing team-based learning (TBL) in accounting courses. Accounting Education, 28(2), 195-219.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. Journal of Economic literature, 47(2), 448-474.

Dearnley, C., Rhodes, C., Roberts, P., Williams, P., & Prenton, S. (2018). Team based learning in nursing and midwifery higher education; a systematic review of the evidence for change. Nurse Education Today, 60, 75-83.

De Santis, A.; K. Sannicandro, K.; Bellini, C.; Cecconi, L. (2019). Valutazione e soft skills nella didattica universitaria, Training actions and evaluation processes, Atti del Convegno Internazionale SIRD, pp. 493-502.

Dohmen, T., Falk A., Huffman D., and Sunde U. (2010). 'Are Risk Aversion and Impatience Related to Cognitive Ability?' American Economic Review 100 (3): 1238–60. https://doi.org/10.1257/aer.100.3.1238.

Espey, M. (2022) 'Variation in Individual Engagement in Team-Based Learning and Final Exam Performance', *International Review of Economics Education*, 41.

Espey, M. (2018). Diversity, effort, and cooperation in team-based learning. *Journal of Economic Education* 49 (1): 8–21.

Espey, M. (2012). 'Enhancing Critical Thinking in Economics Using Team-Based Learning'. https://doi.org/10.22004/AG.ECON.123521.

Griselda, S. (2020). Different Questions, Different Gender Gap: Can the Format of Questions Explain the Gender Gap in Mathematics? Job Market Papers pgr710.

Goh, Cheng Siew, Heap-Yih Chong, Lynne Jack, and Adam Fuad Mohd Faris. (2020). 'Revisiting Triple Bottom Line within the Context of Sustainable Construction: A Systematic Review'. *Journal of Cleaner Production* 252 (April): 119884. https://doi.org/10.1016/j.jclepro.2019.119884.

Good, C., Rattan, A., & Dweck, C. S. (2012). 'Why do women opt out? Sense of belonging and women's representation in mathematics'. *Journal of Personality and Social Psychology, 102*(4), 700–717. https://doi.org/10.1037/a0026659 .

Heckman, J. (1979). 'Sample Selection Bias as a Specification Error', *Econometrica*, Vol. 47, No. 1, 153–161.

Hettler, P. L. (2015). Student demographics and the impact of team-based learning. *International Advances in Economic Research* 21 (4): 413–22.

Imazeki, J. (2015). Getting students to do economics: An introduction to team-based learning. International Advances in Economic Research 21 (4): 399–412.

Jakobsen, K.V., & Knetemann, M. (2017). Putting Structure to Flipped Classrooms Using Team-Based Learning. International Journal of Teaching and Learning in Higher Education 29 (1), 177-185. ISSN: EISSN-1812-9129.

Jann, Ben. «The Blinder–Oaxaca Decomposition for Linear Regression Models». *The Stata Journal: Promoting Communications on Statistics and Stata* 8, n. 4 (2008): 453–79. https://doi.org/10.1177/1536867X0800800401,

Karimi, M., & Biria, R. (2017). Impact of Risk Taking Strategies on Male and Female EFL Learners' Test Performance: The Case of Multiple Choice Questions. Theory and Practice in Language Studies 7 (10), 892-899. doi:10.17507/tpls.0710.10.

Lin, J. W. (2019). The impact of team-based learning on students with different self-regulated learning abilities. Journal of Computer Assisted Learning, 35(6), 758-768.

Michaelsen, L., Neil Davidson, and C. Major. (2014). 'Team-Based Learning Practices and Principles in Comparison with Cooperative Learning and Problem-Based Learning.' *Journal on Excellence in College Teaching*. https://www.semanticscholar.org/paper/Team-Based-Learning-Practices-and-Principles-in-and-Michaelsen-Davidson/cf017faa464c6a7f0ce4707e14242fcbd667959f.

Michaelsen, L. K., and Michael Sweet. (2008). 'The Essential Elements of Team-Based Learning'. *New Directions for Teaching and Learning* 2008 (116): 7–27. https://doi.org/10.1002/tl.330.

Michaelsen, L.K., Knight, A.B., & Fink, L.D. (2004a). Centers for Teaching and Technology. Available at https://bit.ly/3sJzYSQ, last accessed on 22nd April 2021.

Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2004b). *Team-based learning: A transformative use of small groups in higher education*. Sterling, VA: Stylus.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much?. The quarterly journal of economics, 122(3), 1067-1101.

Odell, Kathleen E. 2018. 'Team-Based Learning and Student Performance: Preliminary Evidence from a Principles of Macroeconomics Classroom'. *International Review of Economics Education* 29 (September): 44–58. https://doi.org/10.1016/j.iree.2018.01.001.

Oaxaca, Ronald. 'Male-Female Wage Differentials in Urban Labor Markets'. *International Economic Review* 14, no. 3 (October 1973): 693. https://doi.org/10.2307/255981.

Parappilly, M., Woodman, R.J. & Randhawa, S. (2021) Feasibility and Effectiveness of Different Models of Team-Based Learning Approaches in STEMM-Based Disciplines. Res Sci Educ 51 (Suppl 1), 391–405.

Parmelee, D.X., & Michaelsen, L.K. (2010). Twelve tips for doing effective Team-Based Learning (TBL). Medical Teacher 32 (2), 118-122. doi:10.3109/01421590903548562.

Reardon, S.F., Kalogrides, D., Fahle, E.M., Podolsky, A., & Zárate, R. (2018). The Relationship Between Test Item Format and Gender Achievement Gaps on Math and ELA Tests in Fourth and Eighth Grades. Educational Researcher 47 (5). doi:10.3102/0013189X18762105.

Ruder, P., Maier, M.H. & Simkins, S.P. (2021) 'Getting started with team-based learning (TBL): An introduction', Journal of Economic Education, 52(3), pp. 220–230.

Simkins, S.P., Maier, M.H. and Ruder, P. (2021) 'Team-based learning (TBL): Putting learning sciences research to work in the economics classroom', *Journal of Economic Education*, 52(3), pp. 231–240.

Sparrow, Malcolm K.. (2011) "The GRumbler". Latest update: 11th October 2021 https://scholar.harvard.edu/msparrow/grumbler

Tobin, James. 'Estimation of Relationships for Limited Dependent Variables'. *Econometrica* 26, no. 1 (1958): 24–36. https://doi.org/10.2307/1907382.

Watson, W. E., Michaelsen, L. K., & Sharp, W. (1991). Member competence, group interaction, and group decision making: A longitudinal study. Journal of Applied Psychology, 76(6), 803–809. https://doi.org/10.1037/0021-9010.76.6.803

Nguyen, J.G., Keuseman, K.J., & Humston, J.J. (2020). Minimize Online Cheating for Online Assessments During COVID-19 Pandemic. Journal of Chemical Education 97 (9), 3429-3435. doi:10.1021/acs.jchemed.0c00790.

7. APPENDIX

Table A1 – Description of the main variables

Variable	Name of the variable	Definition
Dependent variable		
OUTCOME IN MACROECONOMICS	*Mark*	Continuous variable which reports the students' verbalized grade in Macroeconomics. It ranges from 18 to 30.
OUTCOME IN MACROECONOMICS	*Pass*	Dummy variable equal to 1 if the student passes the exam and to 0 if he fails.
Independent variables		
EFFECTIVE PARTICIPATION AT TBL	*Treat*	Dummy variable equal to 1 if the student participated in at least 5 over 6 Team-Based Learning lessons and 0 otherwise. *[participation rate higher than 80%]*
FEMALE	*Female*	Dummy variable equal to 1 if the student is a female and equal to 0 if is a male.
PERIOD CONTROL	*Session*	A set of 17 dummy variables which take value 1 in correspondence with one of each 17 different periods 0 for the remaining.
COMPLIANCE AT THE END OF THE FIRST YEAR	*Credits*	Continuous variable equal to the credits that the student earned in the first year. It ranges from 0 to 60 and we considered it important because the macroeconomics course is held in the following year (in the second year).
NATIVE	*Native*	Dummy variable equal to 1 if the student was born in Italy and 0 otherwise.
PREVIOUSLY FACE THE EXAM	*Retaker*	Dummy variable which has a value of 1 if the student is repeating the exam and 0 if the student is Attemptsing the exam for the first time. **variable obtained by manipulating Attempts.*
PERFORMANCE IN HIGHLY QUANTITATIVE SUBJECT	*Highlyquantitative*	Continuous variable which computes the mean[A] of students on exams which have a high quantitative content. By analysing the university courses offering by the department of Economics and filtering out the exams not taken by any of the students in the sample, a pool of 14 exams was created having a highly quantitative content *[see Table A2 for more details]* *[A] mean is computed on the verbalised mark, insufficient marks as well as rejected marks are not detectable.* **Row subgroups means were computed ignoring missing values in the pool.*
PERFORMANCE IN SLIGHTLY QUANTITATIVE SUBJECT	*Slightlyquantitative*	Continuous variable which computes the mean[A] of students on exams which have a medium quantitative content. By analysing the educational offer of the *department* of economics and filtering out the exams not taken by any of the students in the sample, a pool of 7 exams with a slightly lower quantitative content has been detected *[see Table A2 for more details]* *[A] mean is computed on the verbalised mark, insufficient marks, as well as rejected marks, are not detectable.* **Row subgroups means were computed ignoring missing values in the pool.*
PERFORMANCE IN NON QUANTITATIVE SUBJECT	*Nonquantitative*	Continuous variable which computes the mean[A] of students on exams which have not a quantitative content. By analysing the educational offer of the department of economics and filtering out the exams not taken by any of the students in the sample, a pool of 22 exams not having a quantitative content hase been selected *[see Table A2 for more details]* *[A] mean is computed on the verbalised mark, insufficient marks, as well as rejected marks, are not detectable.* **Row subgroups means were computed ignoring missing values in the pool.*
NEIGHBOURHOOD	*NearbyHighSchool*	Dummy variable which has a value of 1 if the student attended a high school in the same region as the university and 0 otherwise.

		**We are aware that the accuracy of the variable is weak for neighbouring regions (as there may be municipalities in other regions closer than those in Emilia Romagna itself), our intention in the medium term is to adjust for the proximity of municipalities even if they are not located in the region.*
UNIVERSITY ENTRANCE SCORE IN READING COMPREHENSION	*ComprehensionAbility*	Continuous variable which reports students' performance in reading comprehension at TOLC. The result is determined by the number of correct (1 point), wrong (-0,25 point) and not given answers (0 points).
UNIVERSITY ENTRANCE SCORE IN MATH	*MathAbility*	Continuous variable which reports students' performance in math at TOLC. The result is determined by the number of correct (1 point), wrong (-0,25 point) and not given answers (0 point).
UNIVERSITY ENTRANCE SCORE IN LOGIC	*LogicAbility*	Continuous variable which reports students' performance in logic at TOLC. The result is determined by the number of correct (1 point), wrong (-0,25 point) and not given answers (0 point).
Minor variables – Used for descriptive statistics or give an in-depth view of the sample		
INTERVENTION DOSAGE	*Dosage*	Continuous variable which ranges from 0 to 6 and considered students' participation at TBL lessons. Dosage was computed by counting (and summing) each Irat score when was not missing. **variables used to generate Dosage originate from six different databases of primary data as a result of downloads from the Moodle platform. The process repeats each year TBL was implemented for a total 18 subdatabase for this sample.*
PARTECIPATION AT TBL	*Participation*	Dummy variable equal to 1 if student participated at least at one Team-Based Learning' lessons and 0 otherwise. *[uneven and fragmented participation rate]* **variables used to generate Participation originate from six different databases of primary data as a result of downloads from the Moodle platform. The process repeats each year TBL was implemented for a total 18 subdatabase for this sample.*
INDIVIDUAL PERFORMANCE AT I-RAT	*IratScore*	Continuous variable which computes the mean of all Irat score collected by students. **variables used to generate IratScore originate from six different databases of primary data as a result of downloads from the Moodle platform. The process repeats each year TBL was implemented for a total 18 subdatabase for this sample.*
EXAM ATTEMPTS NUMBER	*Attempts*	Continuous variable which indicates the number of times the student takes the exam (1 for the first Attempts and progressive number for further tries). **variable generated through lags using session as units of time.*
WINTER SESSION	*Winter*	Dummy variable has a value of 1 if the student takes the exam in the January or February sessions (which are closest to the semester when the macroeconomics course is taught) and 0 if they wait for Summer sessions. **Not in the principal analysis because of collinearity with dummy sessions and because we are aware that is not neutral to TBL (students who participate to TBL tends to cluster themselves in winter session)*
ECONOMIC HARDSHIP INDICATOR	*LowIncome*	Dummy variable equal to 1 if student's family unit has an equivalent economic status indicator lower than 23.000 € and 0 otherwise.
ECONOMIC HARDSHIP INDICATOR	*MiddleIncome*	Dummy variable equal to 1 if student's family unit has an equivalent economic status indicator lower than 45.000 € and 0 otherwise.
ECONOMIC HARDSHIP INDICATOR	*HighIncome*	Dummy variable equal to 1 if the student has not applied for fee reductions. This suggests that student's family unit has an equivalent economic status indicator higher than 45.000 € and 0 otherwise.
TEST ONLINE CISIA	*Tolc*	Continuous variable that reports the students' university entry tests results in which students must solve math, logic and reading comprehension questions.

		The result of each individual test is determined by the number of correct, wrong and not given answers that determine an absolute score, deriving from 1 point for each correct answer, 0 points for each answer not given and a penalty of 0.25 points for each wrong answer. **Not in the principal analysis because it changes across year by the introduction of the English test sub-questions for students who have enrolled from 2017 onwards.*
COMPLIANCE AT THE END OF THE LAST YEAR	*OverTimeGrad*	Dummy variable equal to 1 if the student has not graduated within the prescribed period (April+1 in the last academic year) and 0 otherwise. **Our reference sample does not allow us to use this variable because they are not all students who should already have graduated (the last cohorts of the analysis are students who, in the A.Y. 2019/2020, are in their second year and should have graduated by April 2022. Furthermore, we are not sure to have the updated data even for the A.Y. 2018/2019).*
COMPLIANCE AT THE END OF THE FIRST YEAR	*Credits*	Dummy variable equal to 1 if the student gets at least 40 credits in the first year and 0 otherwise.
DO NOT COMPLETE UNIVERSITY	*DropOut*	Dummy variable equal to 1 if the student left the university without graduating. **As above, but here the time constraint is more relaxed and we expected even less accuracy of the* ***OverTimeGrad*** *variable.*
ENROLMENT WAITING PERIOD	*EnrollGap*	Continuous variable that corresponds to years from graduation to university enrolment.
NEIGHBOURHOOD (NUTS1 ARRANGEMENTS)	*Northeast* *Northwest* *Center* *South&Islands*	4 Dummy variable which takes the value of 1 if the student attended a high school in one of those macroareas (NUTS1) and 0 for the others.

Table A2 – Disaggregation of macro-structures of examination performance

NAME	subject	class
TeachingActivityEC_EIF01	Economics of financial intermediaries	*Highly Quantitative*
TeachingActivityEC_EM01	Monetary economics	*Highly Quantitative*
TeachingActivityEC_FA01	Corporate finance, financial analysis	*Highly Quantitative*
TeachingActivityEC_FA02	Corporate finance	*Highly Quantitative*
TeachingActivityEC_IMI01	Introduction to microeconomics	*Highly Quantitative*
TeachingActivityEC_MA02	Macroeconomics	*Highly Quantitative*
TeachingActivityEC_MEF01	Mathematics for economics and finance	*Highly Quantitative*
TeachingActivityEC_MFA01	Financial and actuarial mathematics	*Highly Quantitative*
TeachingActivityEC_MI03	Microeconomics	*Highly Quantitative*
TeachingActivityEC_MIF01	Models for financial investments	*Highly Quantitative*
TeachingActivityEC_MMF01	Mathematics and financial mathematics	*Highly Quantitative*

TeachingActivityEC_RSFF01	Savings and financial choices of enterprises	*Highly Quantitative*
TeachingActivityEC_SF01	Financial science	*Highly Quantitative*
TeachingActivityEC_ST01	Statistics	*Highly Quantitative*
TeachingActivityEC_EA01	Business economics	*Slightly Quantitative*
TeachingActivityEC_EA02	Business economics 2	*Slightly Quantitative*
TeachingActivityEC_EAC01	Economics of credit companies	*Slightly Quantitative*
TeachingActivityEC_EAC02	Economics of credit companies	*Slightly Quantitative*
TeachingActivityEC_EMM01	Securities market economics	*Slightly Quantitative*
TeachingActivityEC_SW01	Welfare systems	*Slightly Quantitative*
TeachingActivityEC_MA01	Marketing	*Non-quantitative*
TeachingActivityEC_DI01	Industrial law	*Non-quantitative*
TeachingActivityEC_DL02	Labour law	*Non-quantitative*
TeachingActivityEC_DP01	Public law	*Non-quantitative*
TeachingActivityEC_DPC01	Private and commercial law	*Non-quantitative*
TeachingActivityEC_DT01	Tax law	*Non-quantitative*
TeachingActivityEC_DUE01	European Union law	*Non-quantitative*
TeachingActivityEC_EGI01	Economics and business management	*Non-quantitative*
TeachingActivityEC_EI01	International economics	*Non-quantitative*
TeachingActivityEC_EIDI01	Economics and institutions of industrial districts	*Non-quantitative*
TeachingActivityEC_EPL01	Economics and labour policies	*Non-quantitative*
TeachingActivityEC_ERS01	Ethics and corporate social responsibility	*Non-quantitative*
TeachingActivityEC_IEPC01	EU integration and community policies	*Non-quantitative*
TeachingActivityEC_MI01	International marketing I	*Non-quantitative*
TeachingActivityEC_MI02	International marketing II	*Non-quantitative*
TeachingActivityEC_OA01	Business organisation	*Non-quantitative*
TeachingActivityEC_PC01	Programming and control	*Non-quantitative*
TeachingActivityEC_RM01	Marketing research	*Non-quantitative*
TeachingActivityEC_SE01	Economic history	*Non-quantitative*
TeachingActivityEC_SEI	Italian economic history	*Non-quantitative*
TeachingActivityEC_SPE01	Economic history	*Non-quantitative*
TeachingActivityEC_SR	Social responsibility	*Non-quantitative*

Table A3 – Correlation with MARKS and key relationships

	CONTROLS (no TBL)		TREATED (TBL)	
	(1) Marks for Male	(2) Marks for Female	(3) Marks for Male	(4) Marks for Female
Dosage	-0.03	-0.13	-0.01	0.14^{*}
IratScore	-0.04	-0.14^{*}	0.24^{***}	0.30^{***}
Highlyquantitative	0.39^{***}	0.53^{***}	0.48^{***}	0.45^{***}
Slightlyquantitative	0.38^{***}	0.55^{***}	0.41^{***}	0.60^{***}
Nonquantitative	0.48^{***}	0.61^{***}	0.46^{***}	0.43^{***}
TOLC	0.15^{*}	0.33^{***}	0.27^{***}	0.33^{***}
ComprehensionAbility	0.04	0.38^{***}	0.17^{**}	0.09
MathAbility	0.09	0.19^{*}	0.25^{***}	0.40^{***}
LogicAbility	0.10	0.16	0.23^{***}	0.16^{*}
EnglishAbility	0.16	0.04	0.14^{*}	0.15^{*}
Attempts	-0.16^{**}	-0.18^{**}	-0.16^{**}	-0.16^{*}
GapfromDiploma	0.08	0.00	-0.11	0.01
Credits	0.41^{***}	0.58^{***}	0.35^{***}	0.48^{***}

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The entire output of the correlation matrix (all the pairwise correlation coefficients between the 14 variables = a 14×14 matrix for each group) is omitted and only the column concerning the correlations with the dependent variable is reported **(Mark)**).

Table A4 – Goodness of the probit fitting

GLOBAL MODEL

```
              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       406           102  |        508
     -     |        30            47  |         77
-----------+--------------------------+-----------
   Total   |       436           149  |        585

Classified + if predicted Pr(D) >= .5
True D defined as DUMesito != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   93.12%
Specificity                     Pr( -|~D)   31.54%
Positive predictive value       Pr( D| +)   79.92%
Negative predictive value       Pr(~D| -)   61.04%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   68.46%
False - rate for true D         Pr( -| D)    6.88%
False + rate for classified +   Pr(~D| +)   20.08%
False - rate for classified -   Pr( D| -)   38.96%
--------------------------------------------------
Correctly classified                        77.44%
```

FEMALE MODEL

```
              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       162            28  |        190
     -     |        17            36  |         53
-----------+--------------------------+-----------
   Total   |       179            64  |        243

Classified + if predicted Pr(D) >= .5
True D defined as DUMesito != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   90.50%
Specificity                     Pr( -|~D)   56.25%
Positive predictive value       Pr( D| +)   85.26%
Negative predictive value       Pr(~D| -)   67.92%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   43.75%
False - rate for true D         Pr( -| D)    9.50%
False + rate for classified +   Pr(~D| +)   14.74%
False - rate for classified -   Pr( D| -)   32.08%
--------------------------------------------------
Correctly classified                        81.48%
```

MALE MODEL

```
              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       230            52  |        282
     -     |        19            33  |         52
-----------+--------------------------+-----------
   Total   |       249            85  |        334

Classified + if predicted Pr(D) >= .5
True D defined as DUMesito != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   92.37%
Specificity                     Pr( -|~D)   38.82%
Positive predictive value       Pr( D| +)   81.56%
Negative predictive value       Pr(~D| -)   63.46%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   61.18%
False - rate for true D         Pr( -| D)    7.63%
False + rate for classified +   Pr(~D| +)   18.44%
False - rate for classified -   Pr( D| -)   36.54%
--------------------------------------------------
Correctly classified                        78.74%
```

Addendum A5 - Fit test: Simple Tobit VS Craggs Model

Eq 4 ALL $\lambda = 2 * [-262{,}10 + (-814{,}72) - (-1294{,}55)] = 435{,}43$

Eq 4 FEM $\lambda = 2 * [-94{,}49 + (-306{,}8) - (-499{,}51)] = 196{,}44$

Eq 4 MALE $\lambda = 2 * [-145{,}89 + (-472{,}01) - (-765{,}61)] = 245{,}41$

The three values resulting from the above formulae all exceed the chi-square threshold for 30 or 29 degrees of freedom (covariates plus the intercept) of the equations.

Acknowledgments.

We thank the Introductory Macroeconomic Course instructor and tutors, the teaching coordinator of the Department of Economics Marco Biagi and the UNIMORE responsible of the administrative data sets for the assistance provided in administrative data retrieval. We thank an anonymous referee for the stimulating comments received on a previous version of the paper as well as the participants to a Ph.d. Seminar in the Labour, Development and Innovation Phd series of seminars for their insightful comments. The responsibility for the content of the paper and any errors remains exclusively with the authors.

ESSAY N° 4

Gender Wage Gap: within and amongst firms' measurement and determinants.

Chiara Tasselli[a] , Tindara Addabbo[b]

a University of Modena and Reggio Emilia, Marco Biagi Foundation
Email: chiara.tasselli@unimore.it

b University of Modena and Reggio Emilia, Gender Equality Observatory, Marco Biagi Foundation
Address: Viale Berengario 51, 41121, Modena, Italy
Email: tindara.addabbo@unimore.it

Gender Wage Gap: within and amongst firms' measurement and determinants

Abstract: The gender wage gap in Italy has been shown to be persistent and, considering the low and selected participation in the labour market, appears to be much higher than the unadjusted gender wage gap shown by Eurostat data. Data referred to the individual using statistical sources that collect wage and hours of work data (such as ITSILC or SHIW), together with individual and family characteristics, allow to decompose wage differential and correct for the non-random selection of women in the labour market, however, they cannot go deeper into the analysis of the determinants at the firm level. With the intent to unveil the determinants at the firm level, we turned to the analysis of firm microdata that collects, for each firm in the data set, not only microdata on workers' wages and individual characteristics but also the degree of development of policies at the firm level that have been shown to positively contribute to gender equality.

JEL Codes: J31, J71, J81, J16

Key-words: wage gap, Oaxaca decomposition, gender equality

1. Introduction

Extensive research has been done on the gender wage gap in salaries and its negative impact on the status of women in their economic and social lives.

The gender pay gap is a complex issue caused by a combination of interconnected factors such as different allocation of care duties and working time within the household, discontinuous career paths, vertical segregation (sticky floor and glass ceiling), contract types, etc. Discontinuous working profiles are related to the unequal distribution of care work and responsibilities within the couple. Women still bear a higher load of care work and are more likely to interrupt their working profile or reduce the effort in the presence of children.

Lower wages have also been associated with other pre-labour market choices that have been the object of the previous essays in this thesis: the horizontal segregation in education paths which sees females self-clustered in low-remunerative fields. The relevance of this dynamic, in addition to women's lower future professional perspectives, lies in the fact that the choices made by both sexes at a young age are not completely free. They are often biased by society's influences and expectations that relegate women to humanistic careers and see men as more suited to scientific careers (which are characterised by better job opportunities both in terms of salary and employment chances).

Using data from administrative sources of heterogeneous companies that spontaneously submitted themselves to the analysis and certification of gender equality, we provide new empirical evidence on the extent and composition of the gender wage gap. Due to the nature of firms' self-selection, it is realistic to assume that our sample leads to an underestimation of the real wage gap compared to national companies. However, despite the advantageous conditions of the sub-sample, a non-negligible gender pay gap can be observed.

This special type of data source: the treated sample (the entire population within-between enterprises) and variables (presence and trace of organizational policies), allows us to contribute to the existing literature in three ways. First, it allows us to analyze the effects of certain policies and formalizations in companies on the gender pay gap and hypothesize good practices or policies aimed at reducing it. Second, it could initiate a discussion beyond just the gender pay gap, but in terms of retaining female workers within the company and thus reducing self-selection out of the workforce. The Blinder-Oaxaca decomposition allows us to investigate whether this difference is mainly due to gender differences in workers' actual characteristics or whether, instead, it may be due to differential effects of characteristics, perhaps discrimination, and unobserved factors that were not included in the model. Therefore, we can deeply disentangle heterogeneity in firms' pay policies, which results in both direct discrimination and other employer-specific mechanisms driving the gender wage gap as returns in workers’ characteristics. The analysis also allows us to check each variable's weight in explaining the wage differential and the different components - returns and characteristics. Third, this data source can help us understand the role that organizational policies and practices play in creating and perpetuating gender inequality in the workplace, beyond just the gender pay gap. By examining the presence and trace of organizational policies related to issues such as work-life balance, promotion and advancement opportunities, and discrimination prevention, we can gain insights into the factors that contribute to gender disparities in employment outcomes and develop evidence-based recommendations for promoting gender equality in the workplace.

Concerning the latter, special attention is given to horizontal segregation and the relative weight of the gender pay gap to motivate the urgency of orientation policies capable of directing young students towards conscious (and unbiased) educational choices from an early age.

The paper is structured as follows: Section 2 reviews the relevant literature. Section 3 describes the data, data sources and variables, and descriptive statistics for the sample. Section 4 is dedicated to an accurate description of the methodology implemented. Section 5 presents the results of the multivariate analysis and it is divided into two parts: the first (a) using the entire sample aimed at seeing which policies and/or firm characteristics have the greatest impact on women's wages and the second one (b) in which we analyse the subsample of workers (3,126) for whom the level of education is provided. In subsection 6.b , the aim is to reach more precise estimates by controlling for the education level (usually higher for women) and the probability of being employed (higher for more productive women due to self-selection): two components that are very important when estimating wage differentials because both lead to an underestimation of the gap. Section 6 concludes.

2. Literature review

The gender wage gap in Italy has been found to persist over time (Mussida & Piazzalunga 2018; Zizza, 2013) and is higher when one corrects for the non-random selection into employment (Addabbo, 2018; McKay & Mussida 2018; Olivetti and Petrongolo, 2008). The latter is related to women's lower participation in the Italian labour market and differences in the characteristics of employed and not-employed women.

Moreover, by using the Oaxaca-Blinder decomposition, the larger part of it can be attributed to differences in the coefficients rather than to characteristics (Addabbo, 2018, Zizza, 2013).

Quantile regressions disaggregating the population by level of education (Furno 2020, Mussida & Picchio, 2014; Addabbo & Favaro, 2011) detect a higher penalization for lower educated women that increases when one considers their self-selection into the labour market and shows evidence of sticky floor and glass ceiling effects in the change of the estimated wage gap in a different part of the earning distributions and according to the level of education. In the higher level of education, the wage gap at the disadvantage of women increases at the higher quantile of the earnings distribution showing evidence of a glass ceiling effect while in the less educated group, the higher gender wage gap at the bottom of the earnings distribution shows evidence of sticky floor effect.

Also contributing to the gender wage gap is the incidence of household work that is not equally distributed between the gender: Matteazzi & Scherer (2021) founds that partners' housework explains a considerable part of the gender wage gap. Women's housework, in particular, helps men earn more, whereas women seem not to take much advantage in terms of wages from their partners' domestic work.

Wage inequalities are linked also to horizontal and vertical segregation (Blau and Kahn, 2017) .

Centra & Cutillo (2009) address the problem of endogeneity related to the segregation of women into female jobs/sectors by the estimation of a bivariate selection model that controls for the selection in employment and the selection in female occupations on Isfol wage differentials data in Italy. They define female occupations (in terms of sector of employment and profession) as those where the percentage of female workers is higher than 59%. By correcting for the female employment selection, the part of the Oaxaca-Blinder decomposition related to the coefficients of the variables introduced in the model is still higher than the one due to differences in the characteristics but lower than when one does not control for the double selection.

Working in a more 'feminized' workplace has been found to have a negative effect on women's wages also for women in the top positions by Busch and Holst (2011) who estimate models with fixed effects on wage differentials by considering the probability of access to top positions on the German Socio-Economic Panel Study (GSOEP) data.

Busch (2020) finds evidence of an increase in the wage gap between women working in male-typed occupations and women working in other occupations linked to the rise of wages of women working in male occupations occurred and a decline in the discrimination in traditionally male occupations in the last two decades in Germany by using GSOEP data (1992-2015). Addison *et al.*

(2018) use longitudinal Current Population Survey data to analyse the impact of gender composition on earnings and find that its impact is reduced when they control for observed heterogeneity and even more when they control for unobserved heterogeneity and detect when using synthetic panels of ageing cohorts, larger wage penalties for younger cohorts in predominantly female occupations. Low female compensation is an important issue to address as can disincentivize the entry of women into the workforce especially if care duties arises

The increase in the representation of women on the board of public limited liability companies has been found to be related to a decrease in the gender gap in earnings within boards in Norway with no robust impact however on the larger set of women employed in companies subject to the quota (Bertrand et al., 2019).

The importance of analysing the impact on gender wage inequalities of firm-specific pay policies and premiums has been shown amongst others by Card et al. (2016) and by Goldin et al. (2017) showing the need of relying on firm wage data matched with individual data. Card et al. (2016) by using a census of private sector employees by the Portuguese Ministry of Employment matched with longitudinal data on the hourly wages of Portuguese workers, found that firm-specific pay premiums explain about 20% of wage variation by gender, while positive assortative matching explains another 10% and women benefit less from firm-to-firm mobility than men.
With the intent to unveil the determinants at the firm level, we turned to the analysis of firm microdata.

3. Data and descriptive statistics

We use primary data from administrative sources from twelve companies (belonging to sixteen legal entities with 15,168 workers collected from late 2019 to the beginning of 2023, to analyse their gender wage gap. The companies in the sample willingly provided their data by submitting themselves to gender parity analysis and certification and are heterogeneous in terms of size, industry (NACE) and location. The willingness to be evaluated in terms of gender equality can lead to a non-random selection of the observed sample. Another problem with the dataset is the lack of information on education level for more than 75 per cent of the dataset.

On the other hand, as the data collection was finalised to check the gender equality inside the firm, we had a very rich and precise set of variables: we can also control for policies and ratings of the firms in other gender equality dimensions that aren't exclusively related to salaries. In addition, this administrative source allows us to analyze the entire population within-between enterprises. Many

other surveys (such as ITSILC, labour force surveys, etc.) collect randomized sample data and therefore do not have a complete overview at the enterprise level, but at best can trace back to the sectors (NACE), size and geographical areas of enterprises. Our database allows us to better understand the enterprise-worker conditions' interaction.
In the cleaning procedures, to normalize the database, we focus on workers who work between 20[53] to 55 hours per week and have an hourly wage of at least 7 euros[54], omitting 3,056 workers in these two latter categories. Moreover, by following the common trend in the literature which deals with similar datasets we also exclude 171 workers over 65 years old (Blau & Kahn 2017); 183 self-employed[55], 179 collaborators and all the employees for whom it is difficult to trace the hours worked or their income because for them the dependent variable is not observable (Blau & Kahn 2017).

Our main dependent variable is the log of the average hourly earnings (*lnYhour*), which we compute by dividing annual labour earnings by annual hours worked. To have an accurate estimate of the gender pay gap, annual earnings (*Income*) were calculated by adding gross salary to variable revenue and any other elements of compensation. Meanwhile, the yearly hours of work were computed by multiplying the average worker's weekly hours (overtime included) by the number of weeks worked in a year.

Table 1 displays the descriptive statistics for the annual gross earnings of workers (*Income*) the main dependent variable (*lnYhour*[56]) and the covariates used.

We controlled for workers' characteristics such as:

i) Gender (*Female*)
ii) Worker's *age* and Company *seniority*;
iii) Employment contract (*Interim, Apprenticeship, Temporary, Permanent, Shareholder*[57])
iv) Position (*Apprentice, Blue collar, Employee, Middle manager, Manager*)
v) Company area in which they work (Purchase & Logistics "*Purch logistic",* Administration & finance *"Admin fin",* Secretary & services *"Secretary serv", Technical, Commercial,*

[53] Our sample includes a large company operating in the restaurant industry with non-standard workers, whom we have chosen to omit.
[54] As above.
[55] VAT-registered workers.
[56] Further details on the variables and their setting procedure are available in the methodology section.
[57] Shareholders are only 40 individuals and they belong to a cooperative providing integrated service in architecture, engineering and urban planning; we decided to keep them since they had similar characteristics to the typical worker and it was possible to calculate their hourly wages.

Management[58], human resources *"Hr", It, Legal, Marketing, Services*, *Quality safety*, research and development *"Red", Operations.*)

vi) Employee's working pattern: whether he/she is a smart worker (*Smart*) and whether he/she works part-time (*partTime*) in our context defined as those who by contract work no more than 30 hours per week (excluding overtime). We have also included in the descriptive statistics in Table 1 the hours worked per week *(Hweek*) (including overtime)

vii) The level of feminization of the occupation[59] in which the worker is placed (*High_fem*)

viii) Finally, we also use the educational level, for repeating the analysis in the subsample of workers in which the educational level is available (3126). Following the categorization used by the 2021 Labour Force Survey we classify the educational level into 6 classes: Primary education including certificates (*Primary), Middleschool,* partial level completion of High school without direct access to upper secondary education *(Lowhighschool), Highschool,* Master and/[60] bachelor degree *(University),* PhD and Master *(Postgrad).*

Family information (marital status, income earners and members) and motivations for accepting the job are not available. Centra and Cutillo (2009) observe that these do not directly influence wages, but rather the decision to work and the choice of the type of work, which in turn influence wages.

Moreover, we control for some firms' fixed effects (at the legal entity level)[61] as:

i) the *dimension* of the firm in terms of employees (which ranges from 29 to 5,500 employees in our sample)

ii) the logarithmic form of the company revenue (*Ln_revenue*)

iii) the province in which the company is located[62] (*Location*)

Finally, we also add as business fixed effects, other variables which the literature underlines as significant determinants of the gender Pay Gap:

i) share of females on the board (*%F_board*) ;

ii) share of female managers (*%F _Man*)

iii) whether there are formalized parenting policies/services *(RegParents*)

iv) whether there is formalized regulation on hourly flexibility *(RegFlexibility*)

[58] This category is a subgroup of managers: since with this aggregation we are intending to see the sector/area of membership and not the job title; where possible we have included the manager in the area of membership (eg: HR manager in the HR area), but there are still 22 observations left labelled solely with the management.

[59] Occupations were identified based on the intersection of worker's classification and economic activity sector as done by Centra & Cutillo 2009.

[60] unlike the ISCED classification used by Istat, we had to aggregate bachelor's and master's degrees because we did not have the detail for all workers.

[61] Some firms disaggregate into multiple legal entities (we got data from 12 Firms separated into 16 legal entities).

[62] The legal location was considered when the company has more than one physical location.

v) whether there is formalized regulation on part-time *(RegPartTime)*

vi) whether there are policies on professional development *(Profes_Dev)*

As firm fixed effect, these latter are constant variables for workers within the firm and changes between firms and are not displayed in Table 1.

Table 1 - Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Income	11581	30649.46	24886.94	7894.74	519974
Yhour	11581	15.96	11.53	7.01	251.93
Female	11581	.49	.5	0	1
age	11581	46.03	10.32	19.09	65
seniority	11580	9.27	8.75	0	47.49
Primary	3126	.02	.13	0	1
Middle school	3126	.05	.21	0	1
Low high school	3126	.15	.35	0	1
Highschool	3126	.38	.49	0	1
University	3126	.33	.47	0	1
Postgrad	3126	.08	.28	0	1
interim	11581	0	.05	0	1
apprenticeship	11581	0	.06	0	1
temporary	11581	.05	.22	0	1
permanent	11581	.94	.23	0	1
Member	11581	0	.06	0	1
apprentice	11581	0	.05	0	1
blue collar	11581	.60	.49	0	1
white collar	11581	.34	.47	0	1
middle-management	11581	.04	.2	0	1
manager	11581	.02	.12	0	1
Purch logistic	11581	.02	.14	0	1
Admin fin	11581	.05	.21	0	1
Secretary serv	11581	.01	.1	0	1
Technical	11581	.03	.17	0	1
Commercial	11581	.15	.36	0	1
Management	11581	0	.04	0	1
HR	11581	.01	.09	0	1
IT	11581	.01	.08	0	1
Legal	11581	0	.05	0	1
Marketing	11581	.01	.09	0	1
Services	11581	.01	.08	0	1
Quality safety	11581	0	.05	0	1
ReD	11581	.04	.2	0	1
Operations	11581	.67	.47	0	1
smart	11581	.2	.4	0	1
PartTIME	11581	.21	.41	0	1
High_fem	11581	.1	.3	0	1
hweek	11581	35.92	7.07	20	55

The table shows that our final sample consists of 11,581 workers who, on average, have a gross income of 24,886.94, work 35.92 hours per week and earn 15.96 euros per hour. They are about 46 years old and have slightly more than 9 years of company seniority. Nearly half (49%) of the sample is female. Other covariates indicate the share of workers for each position; employment contract; company hours and working pattern. The descriptive statistics report that most workers have a permanent contract (94%) and are in a blue-collar (60%) or white-collar (34%) position. The most populated business area is operation (67%) followed by commercial (15%). It is important to mention that the business areas are strongly influenced by a company in the food service sector, which weighs 55.29 of our sample and whose workforce is almost entirely in this area.

Table 2 allows for analysing of the differences by gender in workers' characteristics and their position within the enterprise. The first block of columns (1) shows the women's average and standard deviation on covariates, the second (2) the same for males and the last (3) presents the results of the statistical test used to compare the averages of two groups in which the null hypothesis (H_0) is that the true difference between these group means is zero. The comparison was made by subtracting the male average ($\bar{x}_m$) from the female average ($\bar{x}_f$) so a positive sign indicates that men's characteristics are higher than women's for that variable and vice versa for the negative sign. As can be seen from the p-value and the t-value there are significant differences (and with the expected signs) for most variables.

Table 2 – Gender comparison on personal characteristics and firm's allocation

	(1) FEMALE		(2) MALE		(3) T-TEST ($\bar{x}_m - \bar{x}_f$)	
	mean	**sd**	**mean**	**sd**	**b**	**t**
Income	24770.22	19177.38	36367.47	28245.86	11597.25***	(25.91)
Yhour	13.98	8.50	17.89	13.57	3.92***	(18.67)
Age	46.79	10.13	45.29	10.45	-1.50***	(-7.86)
Seniority	8.42	7.67	10.10	9.61	1.69***	(10.44)
Primary	0.02	0.15	0.02	0.12	-0.01	(-1.67)
Middleschool	0.07	0.25	0.03	0.18	-0.04***	(-4.13)
Lowhighschool	0.09	0.28	0.18	0.38	0.09***	(7.65)
Highschool	0.32	0.47	0.41	0.49	0.09***	(5.33)
University	0.42	0.49	0.28	0.45	-0.14***	(-7.89)
Postgrad	0.09	0.28	0.08	0.28	-0.00	(-0.21)
Interim	0.00	0.04	0.00	0.06	0.00	(1.55)
Apprenticeship	0.00	0.06	0.01	0.07	0.00	(1.64)
Temporaryfull	0.04	0.20	0.06	0.23	0.01***	(3.52)
Permanent	0.95	0.21	0.93	0.25	-0.02***	(-4.62)
Member	0.00	0.04	0.01	0.07	0.00***	(3.52)
Apprentice	0.00	0.05	0.00	0.04	-0.00	(-1.42)
Blue_collar	0.65	0.48	0.56	0.50	-0.09***	(-9.47)

White-collar	0.31	0.46	0.36	0.48	0.04^{***}	(4.81)
Middle-man.	0.03	0.17	0.06	0.23	0.03^{***}	(8.07)
Manager	0.01	0.09	0.02	0.15	0.01^{***}	(6.44)
Purch_logistic	0.01	0.12	0.03	0.17	0.01^{***}	(5.53)
Admin_fin	0.06	0.23	0.04	0.19	-0.02^{***}	(-4.47)
Secretary_serv	0.01	0.10	0.01	0.10	-0.00	(-0.15)
Technical	0.01	0.10	0.05	0.21	0.04^{***}	(12.10)
Commercial	0.11	0.32	0.18	0.39	0.07^{***}	(10.89)
Management	0.00	0.05	0.00	0.03	-0.00	(-1.76)
Hr	0.01	0.10	0.01	0.08	-0.00	(-1.58)
It	0.00	0.05	0.01	0.09	0.01^{***}	(3.96)
Legal	0.00	0.05	0.00	0.05	-0.00	(-0.81)
Marketing	0.01	0.11	0.01	0.08	-0.00^{**}	(-2.78)
Services	0.01	0.09	0.00	0.07	-0.00^{*}	(-2.49)
Quality_safety	0.00	0.02	0.00	0.06	0.00^{***}	(4.19)
Red	0.03	0.16	0.05	0.23	0.03^{***}	(7.91)
Operations	0.73	0.44	0.60	0.49	-0.13^{***}	(-14.98)
Smart	0.21	0.40	0.19	0.39	-0.01	(-1.79)
Parttime	0.37	0.48	0.05	0.22	-0.32^{***}	(-45.64)
High_fem	0.13	0.34	0.07	0.25	-0.06^{***}	(-11.51)
Hweek	32.69	7.98	39.06	4.09	6.37^{***}	(53.82)
N	**5,710**		**5,871**		**11,581**	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Men earn an average of 11,597.25 euros per year more than women: this is partly due to a greater number of hours worked per week (almost 8 more than women), but we see that a positive difference remains even when we depurate for the number of hours (in fact, men earn 4.09 euros more per hour worked). Even though women in the sample have a slightly higher percentage of permanent contracts they are positioned at the lower end of the grades (blue-collar), while management positions are male-dominated. As many as 37 per cent of women work less than 30 hours a week, while only 5 per cent of men have a part-time job. Moreover, although women are older, they have less seniority: a probable consequence of discontinuous careers due to motherhood and delayed entry into the workforce caused by the higher level of education. In fact, we can see that 42% of women have a university degree compared to only 28% of men and that the difference between the two percentages (14%) is highly significant.

Only the distribution between business areas shows some non-significant differences. The share of workers in different areas is strongly smoothed by the incidence of a big firm which uploads on the operation area. If we removed this company (Table A2 of the appendix) we notice that also horizontal allocation became significant in terms of gender differences with women more concentrated in marketing, secretary, HR, administration and services and males prevailing in IT, operations[63] and logistics.

Table 3 provides an overview of gender differences in earnings and frequencies by business area and respective occupation. Each cell contains the differences in the average of men and women present in that particular business area and position. Since the differences are calculated by subtracting the female mean from the male one, positive numbers indicate female dominance while negative ones have higher values for men.

We can see, in the order, differences in the average income, hourly wages and worker frequency (defined as the number of workers employed in that occupation and for that area). Regarding to the latter it is important to point out that we are looking at absolute values and the gender composition in the sample is slightly unbalanced (5710 Females versus 5871 Males). Nevertheless, this disparity is really minimal: the gap between men and women is only 3% of the numerosity of the two groups when taken separately (171 on 5,710 or 5,871). So, the frequencies are still informative, and moreover, we report a complete overview by gender in Table A3 of the appendix.

By looking at the last row in each cell, we can find further confirmation of unbalanced gender distribution between firms. Row totals confirm the vertical segregation that arose in the previous table by showing that women are underrepresented in management (-86) and middle management positions (-183). This is possible because of a female overrepresentation among blue-collar (+399) and apprentices (+7).

Now also horizontal segregation is more clearly visible in column totals because it does not suffer from the operations predominance. Women are clustered in administration, marketing HR and services. They also have a positive number in the management area, but, as we can see by the cross-tabulations, they are in that area mainly in white collar positions (maybe manager' secretaries) and they earn markedly less than their male counterparts in the same *cell* (area and position).

In this table, we can clearly distinguish how - for almost all cells - regardless of classification or job description women earn significantly less: both on an hourly basis and on an annual basis (the latter also suffers from an even greater gap due to women's shorter working time). The only position where women earn a little more is the apprentice: a low-job position where they are also overrepresented.

With respect to areas -regardless of position (column totals)- women hardly earn more since they are underrepresented in the highest positions (manager and middle manager) of each area. The only

[63] In table 2 the operation area is female-dominated for the particular sector of the highest-weight company: restaurant (catering) and services.

area in which they have higher salaries is the technical one. But this result is mainly driven by the large concentration of men in the blue-collar position for that area (76 vs. 0 for women) rather than by higher women wages.

Table 3 – Horizontal and vertical segregation

AREA	$\overline{\text{female}} - \overline{\text{male}}$ of Income / ln(Your) / freq.					TOTAL
	POSITION					
	Apprentice	Blue_collar	WhiteCollar	MiddleMan	Manager	
Purc_log		-126.84	-7492.24	-11684.69	-124560.3	**-11316.3**
		-0.01	-2.90	-5.66	-60.35	**-0.086**
		-64	-7	-10	-7	**-88**
Admin_fin		-5591.59	-3622.38	-2174.71	5822.25	**-13159**
		-2.53	-0.74	-1.08	2.82	**-0.162**
		-9	122	-6	-13	**94**
Technical			-4934.49	-12716.67		**-5653.79**
			-0.97	-6		**0.045**
		-76	-110	-28	-5	**-219**
Commercial	-1152.25		-2904.66	-8509.59	-10333.98	**-6393.48**
	-0.61		-1.23	-4.17	-5.01	**-0.113**
	3	-9	-347	-67	-19	**-439**
Management			-38423.38		-324048.3	**-206493**
			-16.40		-157	**-1.402**
		-1	8	4	-3	**8**
HR			-4406	-4428.17	-58191.66	**-13782.1**
			-1.60	-2.15	-28.51	**-0.142**
		1	10	2	1	**14**
IT			-2714.43	3013.69		**-8233.64**
			-1.33	1.63		**-0.089**
			-20	-8	-5	**-33**
Legal			691.14	12858.33		**-4076.84**
			0.29	5.47		**-0.049**
			4	1	-1	**4**
Marketing			-2826.18	-7393.34	-15907.08	**-13956.6**
			-1.34	-3.74	-7.71	**-0.214**
		-2	30	-3	2	**27**
Operations	803.79	-8116.79	-11256.09	3256.77	12777.95	**-9268.5**
	0.89	-1.74	-3.92	1.62	6.10	**-0.166**
	-2	597	87	-22	-16	**644**
Secret_serv		-16796.90	-6056.48	-6662.46		**-8817.2**
		-4.22	-1.84	12.69		**-0.139**
	4	-13	11	-1	-1	**0**
Services			-3311.90	2381.89		**-3279.54**
			-1.34	1.46		**-0.086**
			20	1		**21**

Qual_safe			4211.99			**11381.77**
			2.94			**0.412**
		-16	-3	-1	-1	**-21**
ReD	1692.72	211.53	-945.73	-1811.35	-20446.84	**-6346.06**
	0.98	-0.09	0.78	-0.21	-10.20	**-0.072**
	2	-9	-103	-45	-18	**-173**
TOTAL	**493.3**	**-8438.19**	**-5954.85**	**-2766.94**	**-12561.6**	**-11597.3**
	0.028	**-0.151**	**-0.123**	**-0.037**	**-0.064**	**-0.212**
	7	**399**	**-298**	**-183**	**-86**	**-161**

Notes: each cell reports 1) in the first row the differences between females and males income ($\overline{Income_F} - \overline{Income_M}$);
2) in the second row the differences between females and males income ($\overline{\ln\,(Yhour)_F} - \overline{\ln\,(Yhour)_M}$);
3) in the last cell the differences between females and males frequencies ($n°_F - n°_M$).
In cells where only one of the two genders is present, differences in wages it is not calculated.

4. Methodology

As has been found in the literature and shown by descriptive statistics, when examining the gender gap, several variables interfere and looking exclusively at the wage gap could be reductive if not misleading. The starting point for studies of labour income is the Mincerian wage equation (Mincer, 1974), which intends that the impact of certain characteristics, such as level of education, work experience, and age, should be studied in terms of potential productivity. Greater precision of estimation can be achieved by supplementing this equation with additional information about the job of individuals and the characteristics of the firms in which they are placed. Finally, if we incorporate the Mincerian equation with a dichotomous variable indicating gender, we can obtain an initial quantification of the gender wage gap that is not purely descriptive.

First, we look at the wage differential through regressions and afterwards we quantify the discriminative component of that differential using the Oaxaca decomposition. The Oaxaca decomposition is applied to compute how much of this differential is due to observable and objective characteristics and how much, on the other hand, is unexplained and considered a possible consequence of discrimination. In addition, for the subsample of workers for whom we had the level of d education, we also applied the Heckman correction for non-random selection (Heckman 1976, 1979).

In all multivariate analyses, the dependent variable is the hourly wage and appears in logarithmic form: this is common in the literature for two main reasons:

i) it is found that in nature the relationship between wages and productivity proxies is nonlinear;
ii) wages usually have a high skewness due to the presence of few outliers with really high values: applying the logarithmic form allows for greater normalization in the distribution and thus analyses relative (rather than absolute) wages and limits the influence of outliers. This is also

the case of our sample in which we notice heavy outliers and the normality test returned very high values of skewness (p> 0.0000).

The regression in our analyses is structured as reported in equation 1, where subscripts *i* and *f* denote respectively individual and firm, Y_{if} denotes the net hourly wage and X_{if} represents a vector of characteristics of the individual *i* belonging to enterprise *f*, and T_{if} represents a dummy variable indicating if the worker is in a high feminized occupation. Finally, the term ε_{if} , represents the error terms.

The level of feminization of the occupation was created by following the Centra and Cutillo (2009) procedure. We use the data from the fourth quarter 2021 of labour force survey (ISTAT) and we cross-referencing 96 ATECO (Nace two-digit) sectors with gender and 5 worker classification (*Apprentice*, *Blue collar*, *Employee*, *Middle manager, Manager*). This procedure allowed us to compute the share of females in each occupation (intersection of worker's classification and economic activity sector). To control for the "gender clustering" effect in particular occupations, two paths can be followed: we could directly include among the explanatory variables the percentage of women within the particular occupation occupied by the individual; or create a dummy variable, indicative of whether the individual's occupation is a typically female job or not. We chose the second option[64] as done by Centra and Cutillo and identify as threshold the 61% of share of females in the occupation. This value was obtained by multiplying the share of female employment in the entire market (about 40.5 percent) by 1.5, thus inflating the standard line by an additional 50 percent.

$$(1)\quad \log Y_{if} = \alpha + X_{if}\beta_1 + \gamma_f\beta_2 + T_{if}\beta_3 + \varepsilon_{1if} \quad \text{with } i\,(1,\ldots.\,11581)\,;\, f\,(1,16)$$

The full specification of equation 1 is given in equation 5 in the following Section. The equation could be estimated for the whole sample or separately according to gender (as for equations 5a and 5b).

The Oaxaca decomposition is obtained by subtracting from the male wage function, the female one (the estimate is computed at the mean values of the variables). Equation 2 reports the formula in its reduced form[65]:

$$(2)\quad \overline{\Delta \log Y} = (\hat{\alpha}_m - \hat{\alpha}_f) + (\hat{\beta}_m - \hat{\beta}_f)\cdot \bar{X}_f + \hat{\beta}_m \cdot (\bar{X}_m - \bar{X}_f) + \varepsilon$$

[64] However, all the analyses were repeated with the continuous variables and the coefficients remain stable in both sign and significance.

[65] To simplify notation, all characteristic vectors of individuals, firms, and years are resumed in X as they may well all be processed as individual identifiers. Subscripts have also been removed to focus on the algebraic gender designation, but initial specification remain implicit.

where the dash *($\bar{X}$)* indicates the vector of mean values of the characteristics used, the hat ($\hat{\beta}$) indicates the vector of estimated coefficients, and the indices M and F indicate, respectively, the collective of men and the collective of women.

The first part of equation 2 is named the discriminatory component: difference in the wages intercept $(\hat{\alpha}_m - \hat{\alpha}_f)$ corresponds to direct discrimination (gap due exclusively to gender), meanwhile $(\hat{\beta}_m - \hat{\beta}_f) \cdot \bar{X}_f$ relates to the possibility of men receiving higher pay for their characteristics.

Finally, the latter part of the equation $\hat{\beta}_m \cdot (\bar{X}_m - \bar{X}_f)$ is called the “endowment effect” and is not considered to be discriminatory because it computes the component of the wage gap arising from actual differences in characteristics between genders (as education, seniority, contract, company position, …). Even on this component, however, a reference should be made to discriminatory phenomena. As mentioned in the previous sections, differences in characteristics could also result from women's barriers to access and accessibility of senior roles (glass ceilings) or the difficulty of changing their status if they enter low positions (sticky floor). Alongside vertical segregation, horizontal segregation is also consistent in the labour market: however, if the former is purely attributable to employer discrimination, the latter is strongly dependent on the educational choices[66] of women (*pre-labour market discrimination*) who tend to specialize in lower-paying fields.

Last but not least: it is worth mentioning that this decomposition is, however, applied to a sub-sample of workers that is not randomly selected: by analysing data on workers we are omitting all those workers who are not employed. The greater the divergence between the population and the subsample of workers, the greater the probability of heterogeneity between the two groups in observed and unobserved characteristics due to self-selection. This is particularly relevant in Italy where the lower number of employed women is matched with high heterogeneity between women who are employed and those who are not employed. In fact, Istat data on the labour force show that only 53.2%[67] of women between 20 and 64 years are employed, and as anticipated in the literature section these are the most educated ones who can earn a higher wage than their reservation wage.

For that reason, in labour market research, it is common to include a correction for sample selection in the wage equations based on Heckman's (1976, 1979) procedure.

[66] The literature and empirical studies -however- reveal how these choices are influenced by the stereotyped view of society. Stereotypes of teachers (Carlana 2019; Lavy 2008), parents (Carlana and Corno 2021) and peer behaviours (Astorne-Figari and Speer 2019; Booth et al. 2018) can strongly affect girls' choices and performance in quantitative studies.

[67] employment rate in 2021, this value falls to 35.7 per cent in the South of Italy.

From an econometric point of view, this assumption means that our dependent variable in *equation 1* is not always observed. Rather, the dependent variable for observation j is observed if:

$$(3)\ z_{if}\omega + \varepsilon_{2if} > 0$$

where

$$\varepsilon_{1if} \sim N(0; \sigma)$$

$$\varepsilon_{2if} \sim N(0; 1)$$

$$corr(\varepsilon_{1if}; \varepsilon_{2if}) = \rho$$

z_j is a vector of variables that strongly affect the chances to have an observable dependent variable[68] (the reservation wage) but not the outcome under study (the offered wage). We select variables such as educational level, age and the number of children.

The main issue is the correlation between the two error terms ($\rho \neq 0$) that lead to biased results for *equation 1*, and we can correct it by integrating our equation with the fitted values of the selection equation which are in a function called the inverse Mill's ratio (Verbeek, 2014). We calculate the Mill's ratio by merging our dataset with the fourth quarter 2021 of the Italian labour force survey data and estimate a Probit model on the women's probability of being employed. The latter was calculated using employed status as the dichotomous dependent variable (1= employed, 0 otherwise) and age, educational qualification, and the number of children as explanatory variables. The variables were all standardized and homogenized[69] to the same classes and units of measurement and the ISTAT sample was circumscribed to only 20-64 aged women aged, excluding students and not vat holders or collaborators. The coefficients ($\widehat{xb}$) of our estimates were stored and used for computing the Mill's ratio estimates by following the Jann (2008) procedure as for equation 4.

$$\textbf{(4)}\ mills = \frac{\varphi(-xb)}{(1-\Phi(xb))}$$

Where φ is the density function of the standardized normal distribution and Φ is the distribution cumulative standardized normal. Finally, as anticipated, we entered the Mill's ratio in women's wages equation and compute the corrected Oaxaca estimates.

[68] For the wage to be observable it is necessary to be employed.

[69] One obstacle we faced was that our database did not have variables related to the family caregiving burden, and we try to overcome it by using the number of dependent children as a proxy.

5. Multivariate analysis Results and Policies implications

a. Multivariate analysis for all the sample

The gender wage gap is defined by Eurostat as "the difference in average gross hourly earnings between women and men employed in firms with more than 10 employees" and is expressed as a percentage of the corresponding male wage. In our sample, by following the Eurostat definition, women earn only 78% of male hourly wages. The average hourly income for women is €13.98/hr, while for the male counterpart is € 17.89/hr: With a difference of €3.91/hr, which according to the t-test has a high level of significance. From this initial exploratory analysis, it appears that there are gender wage differences (both considering time spent working, but also considering equal hours worked). The wage gap just calculated, however, is a raw statistic and does not take into account many other factors that may be simultaneously correlated with gender and income level.

In practice, the gender variable (*Female*), alone, is only able to explain around 6.7% of the variation in the hourly wage. In order to improve our estimates, it is preferable to focus on wage discrimination (rather than the wage gap). Wage discrimination occurs when a gender gap in gross wages persists even when the comparison is made on a par with any other condition (Becker 1971). If we also include age, for example, we see that the two variables together can explain about 8.1% of the variation in hourly wages. But it is important to note that if instead of age, we enter the hours worked per week we succeed in explaining the 18.2% of the variation in hourly wages. This means that the time dedicated to work can explain variations in hourly salary more than twice the worker's age. The sign of the coefficient of weekly hours is positive and highly significant, which is not obvious since by having hourly wages as the dependent variable we are already cleaning up economic income by the number of hours worked. If the employees' number of hours became an important variable in explaining salary we are facing indirect gender discrimination: this happens because females are strongly associated with fewer working hours whereby they would be the most affected by this dynamic.

We introduce further reasoning and analysis of the correlations between time and hourly wages in Figures 1 and 2.

FIGURE 1 - Prediction of weekly hours worked and hourly wage - a breakdown of age by gender

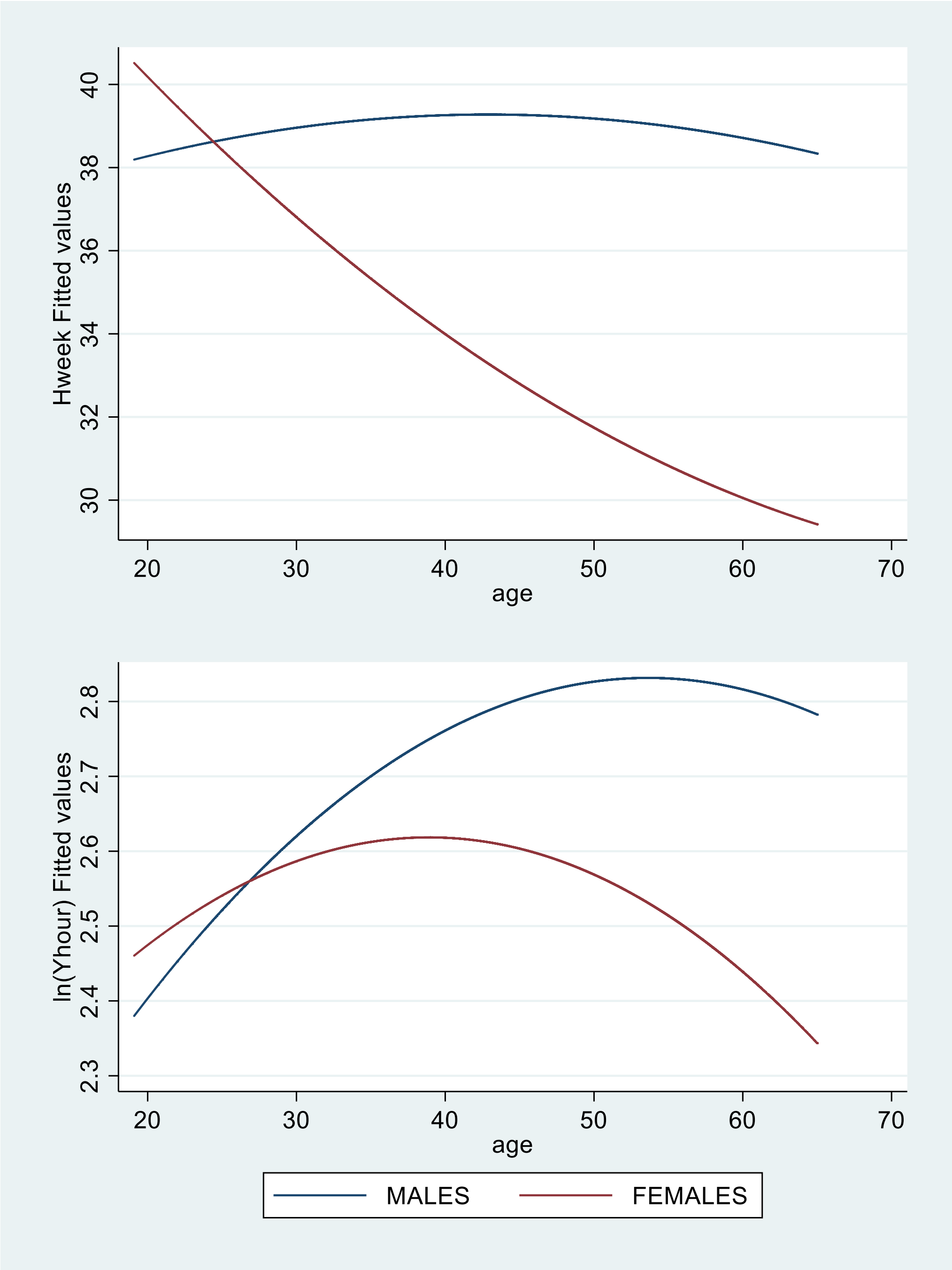
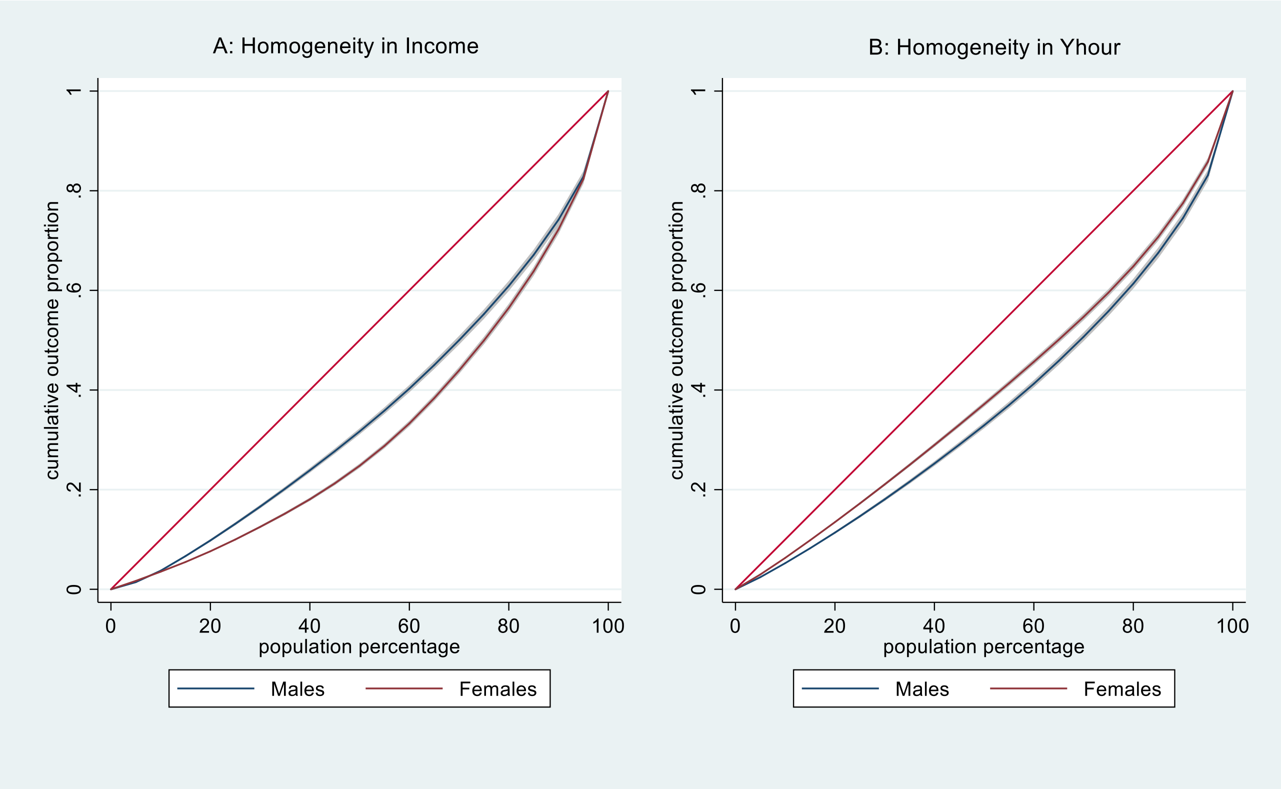


Figure 1 reports the fitted values of the hourly week (panel A above) and the logarithmic form of the hourly wage (panel B below) regressed on age (horizontal axes) by gender.

The part above shows that while in this sample men show a parabolic trend in the hours they spend at work (spending more time at work in their 35 to 50 years); it is noticeable that women spend less and less time at work even though at the beginning of their careers, they work more than men. If we repeat the analysis by replacing the dependent variable with the hourly wage (below in the figure) the trend is expected: we see that shortly after the decrease in hours worked the hourly wage also begins to decrease. Around age 25 we see a reversal in hours worked between men and women and a few years later (approximately 27) also in hourly wages. Here it is important to remember what has been emphasized since the beginning of the analysis: namely, that by using hourly wages as a dependent variable we should have already adjusted earnings from hours worked.

Figure 2(a) and (b) – Gini and Lorenz curve of the earnings distribution by gender



*Notes: Gini associates with the Lorenz curve are the following:

A (Income)	Female' Gini: 0.322	Males' Gini:0.279
B (Yhour)	Female' Gini: 0.218	Males' Gini:0.265

Figure 2 reports in part A the females and males[70] Lorenz curve for the annual gross salary. Regardless of whether males have important outliers (observation with extremely high salaries) in income, the female population has less wage equidistribution: this result was unexpected. We could find an explanation in the heterogeneity of women's labour supply (number of hours worked per week). In fact, by repeating the analysis with the hourly wage (part B - which adjusts for the different labour supply) we see an inversion of the positions of the Lorenz curves.

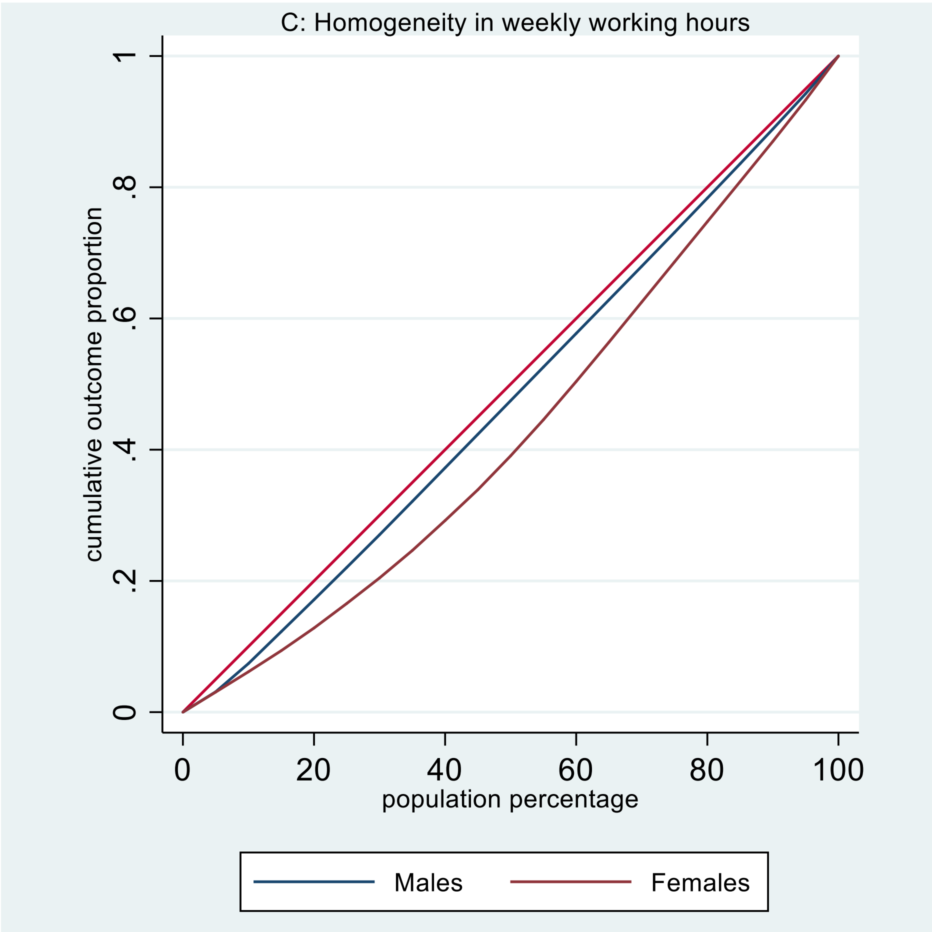
In fact, while for men the Gini coefficient remains basically stable between income (0.28) and hourly wages (0.27), the women one goes from 0.32 on annual income (greater non homogeneity due to different hours worked) to a higher level of homogeneity (0.22) when looking at hourly wages (flatter female wages). Women's hourly wages are more likely to be homogeneous because of their difficulty in reaching top positions (glass ceiling and sticky floor), in fact we can see in our

[70] We can see that male's confidence interval is thicker (probably because of outliers due to high income).

database a huge gender difference in the maximum value for an hourly wage in the sample (€251.92hr for males vs €192.90hr) rather than the minimum one that it is pretty similar.

The relationship between parts A and B and a further confirmation of our suppositions in figure 3 in which the Lorenz curves for the weekly working hours are represented. Men have almost perfect equidistribution of the hours they spend working per week; whereas women, show a greater variability due to downward variations in work time (part-time).

Figure 2 (c) – Gini and Lorenz curve of weekly hours worked by gender



*Notes: Gini associates with the Lorenz curve are the following:

C (Hwek) Female' Gini: 0.137 Males' Gini:0.042

Table 4 shows the estimation of equation 1 in the methodological section, more precisely in its full specification we can decline it as follows (eq. 5) and replicate it for the two gender subsamples (eq. 5a and 5b).

$$ln(Y_{hour}) = \alpha_0 + \beta_1 Female + \beta_2 Age + \beta_3 Age^2 + \beta_4 Seniority + \beta_5 Seniority^2 + \beta_6 Interim + \beta_7 Apprenticeship + \beta_8 Permanent + \beta_9 Member + \beta_{10} BlueCollar + \beta_{11} WhiteCollar + \beta_{12} MiddleManager + \beta_{13} Manager + \beta_{14} Smart + \beta_{15} PartTime + D_1 Area + \beta_{16} HigFem + \beta_{17} \%Fboard + \beta_{18} \%FManager + \beta_{19} RegParents + \beta_{20} RegFlexibility + \beta_{21} RegParTime + \beta_{22} ProfesDev + \beta_{23} Dimension + \beta_{24} LnRevenue + \varepsilon_1 \quad (5)$$

$$\ln(Y_{hour})|_{female} = \alpha_0 + \beta_1 Age + \beta_2 Age^2 + \beta_3 Seniority + \beta_4 Seniority^2 + \beta_5 Interim + \beta_6 Apprenticeship + \beta_7 Permanent + \beta_8 Member + \beta_9 BlueCollar + \beta_{10} WhiteCollar + \beta_{11} MiddleManager + \beta_{12} Manager + \beta_{13} Smart + \beta_{14} PartTime + D_1 Area + \beta_{15} HigFem + \beta_{16}\%Fboard + \beta_{17}\%FManager + \beta_{18} RegParents + \beta_{19} RegFlexibility + \beta_{20} RegParTime + \beta_{21} ProfesDev + \beta_{22} Dimension + \beta_{23} LnRevenue + \varepsilon_1$$

$$\ln(Y_{hour})|_{male} = +\beta_1 Age + \beta_2 Age^2 + \beta_3 Seniority + \beta_4 Seniority^2 + \beta_5 Interim + \beta_6 Apprenticeship + \beta_7 Permanent + \beta_8 Member + \beta_9 BlueCollar + \beta_{10} WhiteCollar + \beta_{11} MiddleManager + \beta_{12} Manager + \beta_{13} Smart + \beta_{14} PartTime + D_1 Area + \beta_{15} HigFem + \beta_{16}\%Fboard + \beta_{17}\%FManager + \beta_{18} RegParents + \beta_{19} RegFlexibility + \beta_{20} RegParTime + \beta_{21} ProfesDev + \beta_{22} Dimension + \beta_{23} LnRevenue + \varepsilon_1$$

Where the D_1 is a set of dummy variables for the 13 areas.

The regressions, in addition to being repeated by gender, are replicated according to the number of hours worked: the first group refers to the workers having a contract of at least 20 hours per week (column 1 all the sample, column 2 the male subsample and column 3 the female ones) and the second group (column 4 all, 5 males and 6 females) those who work at least 30 hours.

The adjusted R-square reports a very good fit of the model indicating that the variables used can explain about 80% of the variations in hourly wages.

Regardless of the weekly hours worked, the gender coefficient is negative and statistically significant (p-value < 0.01), age and seniority have a positive impact on wages: age seems to be more important for the subsample of the workers who work at least 30 hours per week, meanwhile, seniority has a greater impact for females only if they work more than 30h per week. Having a permanent contract has a significant and positive influence on hourly wages, especially for females and both coefficient and significance increase with the increase in weekly working hours. Members (especially if males) have lower hourly wages, but it may depend on the redistribution criteria: they generally have lower hourly wages compared to employees of traditional companies because their main objective is not to maximize the profits of the company, but to pursue the common welfare of the members through the production of goods or services useful to the community. As their goal could be to maximize long-term value for company members, rather than focusing on maximizing profit for shareholders. This can lead to lower hourly wages for members compared to employees of traditional companies, but it can also lead to greater stability and job security. The dummy variable related to smart working is associated with higher salaries this could be because the types of work and tasks that can also be done in smart working usually start at a medium-high level (no manual, blue collar or front office tasks).

As expected, the working position (base apprentice) shows significantly larger coefficients than the previous variables. It should be noted, however, the coefficients for women are smaller than for men (especially for higher levels as managers and middle managers). Part-time, on the contrary, have a negative effect on hourly wages, and this is in line with what was previously commented on in figure 2. The effect seems to be greater for males it should be remembered that very few males are affected by this status. The dummy variable *High_fem* report that being in a highly feminized occupation results in significantly lower wages (especially for women) and this is in line with previous studies (Centra and Cutillo 2009; Fort & Schneeweis 2013; Pastore & Verashchagina, A. 2015) and the *"Annual Labour Force Survey - 2020 Results"* published by Istat in September 2021.

Turning to the firms' fixed effect we notice that the higher percentage of females in the managerial position *(%F_Man*) is related to higher wages (both for men and for women and especially if they work more than 30 hours per week). Meanwhile, the share of females on the board *(%F_Board*) has a negative relation with salaries, especially for males. It's unlikely that an increase in the share of females on corporate boards would directly lead to lower salaries. In fact, having a higher share of women on corporate boards has been associated with better financial performance and corporate governance (Del Giudice, Scuotto & Carayannis 2020; Campa & Quaglione, 2019; Gregorič, & Cernoša 2019), as well as increased gender diversity and representation in leadership positions.

However, other factors may be related to both the gender composition of corporate boards and the wages in Italy. For example, there is a legal requirement for Italian companies to have a minimum number of female board members, which applies only to some types of company. This requirement was introduced in 2011 and applies to all companies listed on the Italian stock exchange and state-controlled companies.

Table 4 – Wage Equations by gender and hours of work

	Al workers (At least 20h)			At least 30h		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	M	F	All	M	F
Female	-0.078***			-0.086***		
	(0.004)			(0.004)		
Age	0.021***	0.021***	0.021***	0.024***	0.024***	0.026***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Agesq	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Seniority	0.006***	0.007***	0.006***	0.006***	0.005***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Senioritysq	-0.000**	-0.000*	-0.000*	-0.000*	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Interim	-0.061	-0.100*	-0.026	-0.051	-0.084	-0.022
	(0.034)	(0.045)	(0.052)	(0.036)	(0.045)	(0.057)
Apprenticeship	0.056*	0.025	0.041	0.065*	0.038	0.050
	(0.028)	(0.038)	(0.041)	(0.030)	(0.038)	(0.047)
Permanent	0.023*	-0.012	0.034**	0.038***	0.005	0.049***
	(0.009)	(0.014)	(0.013)	(0.010)	(0.014)	(0.015)
Member	-0.162***	-0.180**	-0.143*	-0.133**	-0.169**	-0.090
	(0.041)	(0.055)	(0.069)	(0.042)	(0.055)	(0.076)
Blue_collar	0.168***	0.212***	0.122**	0.143***	0.192**	0.098*
	(0.037)	(0.063)	(0.043)	(0.038)	(0.062)	(0.048)
WhiteCollar	0.461***	0.487***	0.450***	0.434***	0.470***	0.406***

	(0.036)	(0.063)	(0.042)	(0.038)	(0.062)	(0.047)
MiddleManager	1.001***	1.031***	0.989***	0.975***	1.014***	0.951***
	(0.037)	(0.063)	(0.044)	(0.039)	(0.063)	(0.049)
Manager	1.706***	1.738***	1.667***	1.688***	1.727***	1.632***
	(0.039)	(0.065)	(0.050)	(0.040)	(0.064)	(0.055)
Smart	0.100***	0.095***	0.094***	0.083***	0.081***	0.071***
	(0.007)	(0.010)	(0.010)	(0.007)	(0.010)	(0.011)
Parttime	-0.122***	-0.177***	-0.106***			
	(0.005)	(0.012)	(0.005)			
Purch_logistic	-0.033*	-0.025	-0.051*	-0.031*	-0.023	-0.045
	(0.013)	(0.016)	(0.021)	(0.013)	(0.016)	(0.024)
Admin_fin	0.071***	0.020	0.092***	0.048***	0.002	0.075***
	(0.011)	(0.016)	(0.015)	(0.011)	(0.016)	(0.017)
Secretary_serv	0.097***	0.104***	0.071**	0.088***	0.093***	0.072*
	(0.018)	(0.027)	(0.026)	(0.020)	(0.027)	(0.031)
Technical	0.189***	0.186***	0.144***	0.173***	0.176***	0.130***
	(0.015)	(0.020)	(0.028)	(0.016)	(0.020)	(0.032)
Commercial	0.031***	0.023*	0.045***	0.023**	0.018	0.039**
	(0.008)	(0.011)	(0.013)	(0.009)	(0.011)	(0.014)
Management	0.156***	0.665***	-0.077	0.173***	0.650***	-0.059
	(0.039)	(0.074)	(0.044)	(0.042)	(0.073)	(0.050)
Hr	0.049*	0.057	0.045	0.059**	0.062*	0.060*
	(0.020)	(0.032)	(0.025)	(0.021)	(0.032)	(0.028)
It	0.103***	0.063*	0.177***	0.094***	0.054	0.176***
	(0.024)	(0.030)	(0.042)	(0.025)	(0.029)	(0.046)
Legal	0.069*	0.029	0.105*	0.073*	0.028	0.117*
	(0.034)	(0.055)	(0.042)	(0.036)	(0.054)	(0.048)
Marketing	0.099***	0.111***	0.093***	0.102***	0.113***	0.098***
	(0.019)	(0.033)	(0.024)	(0.020)	(0.032)	(0.026)
Services	-0.128***	-0.182***	-0.086**	-0.145***	-0.194***	-0.099**
	(0.023)	(0.039)	(0.027)	(0.024)	(0.039)	(0.030)
Quality_safety	0.144***	0.113**	0.470***	0.143***	0.114**	0.442***
	(0.037)	(0.041)	(0.118)	(0.038)	(0.041)	(0.130)
Red	0.292***	0.230***	0.351***	0.216***	0.158***	0.286***
	(0.015)	(0.024)	(0.022)	(0.017)	(0.024)	(0.026)
High_FEM	-0.116***	-0.054***	-0.171***	-0.116***	-0.056***	-0.161***
	(0.009)	(0.014)	(0.013)	(0.009)	(0.014)	(0.015)
%F_board	-0.249***	-0.349***	-0.212***	-0.440***	-0.496***	-0.374***
	(0.043)	(0.076)	(0.054)	(0.054)	(0.080)	(0.073)
%F_Man	0.243***	0.266***	0.258***	0.296***	0.331***	0.265***
	(0.038)	(0.061)	(0.051)	(0.041)	(0.061)	(0.058)
RegParents	0.025***	0.033***	0.015***	0.031***	0.037***	0.024***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
RegFlexibility	-0.103***	-0.151***	-0.047*	-0.123***	-0.155***	-0.085**
	(0.017)	(0.025)	(0.024)	(0.018)	(0.025)	(0.028)
RegPartTime	0.003	0.013	0.000	0.001	0.005	0.010
	(0.006)	(0.008)	(0.009)	(0.006)	(0.008)	(0.011)
Profes_Dev	-0.130***	-0.144***	-0.093***	-0.123***	-0.141***	-0.079***
	(0.010)	(0.017)	(0.013)	(0.012)	(0.018)	(0.017)
Dimension	0.000***	0.000	0.000**	0.000**	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln_revenue	0.007	-0.015	0.016**	-0.026***	-0.043***	-0.013

	(0.005)	(0.010)	(0.005)	(0.007)	(0.011)	(0.009)
Constant	1.766***	2.227***	1.489***	2.354***	2.691***	1.970***
	(0.101)	(0.189)	(0.115)	(0.137)	(0.204)	(0.183)
N	11580	5870	5710	9175	5575	3600
Adjusted R^2	0.803	0.796	0.793	0.785	0.793	0.773

Notes:
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Temporary variable omitted for collinearity in contract (base for *interim; apprenticeship; permanent; Member*)
Apprentice variable omitted for collinearity in position (base for *blue_collar; WhiteCollar; MiddleManager; Manager*)
Operation variable omitted for collinearity in areas (base for *Purchase_logistic Administration_finance Technical Commercial Management HR IT Legal Marketing Secretary_services Services Quality_safety ReD*)

Table 5, shows the results for the Oaxaca decomposition presented in equation 2 of the methodology section. The decomposition is applied by entering in equation 2 the same combination of covariates shown in equations 5a and 5b. As reported in equation 2, the Oaxaca decomposition is obtained by subtracting from the male wage function, the female one, so negative values in table 5 mean higher characteristics and/or coefficients for women, while positive values are in favour of men.

Here the most important evidence that arose from different subsamples is that when we reduce the sample by cutting those who have fewer hours of work per week the gender gap due to explained characteristics suffers a huge decrease. In fact, the part-time coefficient in the explained part is the one with a higher magnitude across all regressors. This is a piece of important evidence and is consistent with the indirect discrimination rationales we raised by analysing Figures 1 and 2.

The difference in earnings is in favour of men and amounts to 21.2% of hourly wages (13.4% of the component explained by characteristics and 7.8% the unexplained component) for the sample which includes 20h workers. The same gap is halved at 10,8% when we analyse those who work more than 30h. Differences in characteristics decrease at 2.3%, meanwhile, the unexplained part slightly increases at 8.6%.

The explained part in estimates (1) reports results extremely in line with descriptive statistics. Although women are older they have lower firm seniority (possible causes could be higher education and discontinuous careers). They are under-represented in higher positions (manager and middle manager). The positive sign on part-time characteristics means that this variable is in favour of males. We could interpret this result by assuming that women are the higher part-time users and part-time is a characteristic associated with lower salaries, so the part-time characteristic is in favour of males for their lower take up. The same interpretation can be provided to the high-feminization occupations having a positive and significant coefficient. As regards the unexplained part we see that having a permanent contract or a part-time contract is more remunerative for females meanwhile males in highly feminized occupations earn on average 1.2% more than females in the same occupations. This suggests that there may be other factors, such as discrimination or differences in negotiation skills, that are contributing to the gender wage gap in these occupations. The estimates for the subsamples who work more than 30h (2) are pretty similar, the only difference that we see is that age is no more a characteristic associated with higher female wages, but on the other side, being a smart worker does.

When we move to the firm fixed effect (below in the table) we notice that being in a firm with a high share of women on the board has a statistically significant positive impact on females' wages. As regards the impact of the share of female managers has a practically negligible effect on the gender pay gap, but the regressions in Table 4 suggested that it had a positive influence on both male and female salaries.

Interestingly note that having formalized regulatory systems have different impacts on female wages: in fact, formalized policies/services dedicated to parenting negatively affect females' wages while formalized hourly flexibility is a firm's characteristic which increases women's wages. The negative effect of formalized policies/services dedicated to parenting on females' wages may not necessarily be associated with a negative connotation: it could be, for example, that allowing people to devote themselves to their personal lives (or care duties) reduces the probability of quitting their jobs: companies with more formalized policies/services dedicated to parenting tend to have lower female wages, perhaps because they can retain mothers (characterized by lower workload) and have lower self-selection of female workers. In fact, the magnitude of the coefficient for this variable is greater for the subsample which includes the 20h workers.

While having formalized hourly flexibility is a firm characteristic which has an unquestionable positive effect on women's wages as could enable women's work without reducing their working time or their performance. Concerning the presence of a policy of professional development seems to be a firm characteristic in favour of women's wages for 20h workers and which has higher returns for female workers >30h.

Finally, firms' size (in terms of employees) and revenues seem to be factors which have a positive impact on women if we consider the 20h sample.

Table 5 – Oaxaca Decomposition

	(1) All workers (At least 20h)			(2) At least 30h		
	Overall decomposition	**Explained**	**Unexplained**	**Overall decomposition**	**Explained**	**Unexplained**
Male	2.764***			2.787***		
	(0.006)			(0.006)		
Female	2.552***			2.678***		

	(0.005)			(0.006)		
Difference	0.212***			0.108***		
	(0.007)			(0.008)		
Explained	0.134***			0.023**		
	(0.007)			(0.008)		
Unexplained	0.078***			0.086***		
	(0.004)			(0.004)		
Age		-0.032***	-0.001		0.008	-0.116
		(0.005)	(0.128)		(0.005)	(0.143)
Agesq		0.026***	0.047		-0.007	0.120
		(0.004)	(0.068)		(0.005)	(0.077)
Seniority		0.010***	0.007		0.010***	-0.009
		(0.001)	(0.012)		(0.002)	(0.015)
Senioritysq		-0.003*	-0.000		-0.003*	0.006
		(0.001)	(0.007)		(0.002)	(0.008)
Interim		-0.000	-0.000		-0.000	-0.000
		(0.000)	(0.000)		(0.000)	(0.000)
Apprenticeship		0.000	-0.000		0.000	-0.000
		(0.000)	(0.000)		(0.000)	(0.000)
Permanent		-0.000*	-0.044*		0.000	-0.041*
		(0.000)	(0.020)		(0.000)	(0.018)
Member		-0.001*	-0.000		-0.000	-0.000
		(0.000)	(0.000)		(0.000)	(0.001)
Blue_collar		-0.014***	0.054		0.008***	0.048*
		(0.003)	(0.033)		(0.002)	(0.024)
WhiteCollar		0.019***	0.012		-0.037***	0.026
		(0.004)	(0.018)		(0.005)	(0.019)
MiddleManager		0.030***	0.002		0.017***	0.003
		(0.004)	(0.002)		(0.005)	(0.003)
Manager		0.025***	0.001		0.019***	0.002
		(0.004)	(0.001)		(0.005)	(0.001)
Smart		-0.001	0.000		-0.009***	0.003
		(0.001)	(0.004)		(0.001)	(0.005)
Parttime		0.039***	-0.009***			
		(0.002)	(0.002)			
Purch_logistic		-0.002***	0.000		-0.001*	0.000
		(0.000)	(0.001)		(0.000)	(0.001)
Admin_fin		0.000	-0.003*		0.001	-0.003*
		(0.000)	(0.001)		(0.001)	(0.002)
Secretary_serv		-0.000	0.000		-0.000	0.000
		(0.000)	(0.000)		(0.000)	(0.001)
Technical		0.004***	-0.000		0.003***	0.000
		(0.001)	(0.001)		(0.001)	(0.002)
Commercial		-0.004***	-0.005		-0.001	-0.004
		(0.001)	(0.003)		(0.000)	(0.004)
Management		-0.000	0.001**		-0.000	0.002**
		(0.000)	(0.000)		(0.000)	(0.001)
Hr		0.000	0.000		0.000	0.000
		(0.000)	(0.000)		(0.000)	(0.001)
It		0.000	-0.001**		0.000	-0.001**
		(0.000)	(0.000)		(0.000)	(0.000)

Legal	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Marketing	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.001)
Services	0.001*	-0.001*	0.002***	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
Quality_safety	0.000	-0.000**	0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Red	0.006***	-0.005**	0.003***	-0.006**
	(0.001)	(0.002)	(0.001)	(0.002)
Operations	0.011***	0.002	0.001	0.002
	(0.002)	(0.016)	(0.001)	(0.014)
High_FEM	0.007***	0.012***	0.013***	0.013***
	(0.001)	(0.003)	(0.002)	(0.004)
%F_board	-0.006***	-0.018	-0.008***	-0.016
	(0.001)	(0.014)	(0.001)	(0.016)
%F_Man	0.002*	0.002	-0.002*	0.012
	(0.001)	(0.019)	(0.001)	(0.020)
RegParents	0.039***	0.073***	0.015***	0.060*
	(0.005)	(0.021)	(0.004)	(0.025)
RegFlexibility	-0.015***	-0.131*	-0.007***	-0.092
	(0.004)	(0.065)	(0.002)	(0.068)
RegPart	-0.001	0.009	-0.000	-0.004
	(0.002)	(0.012)	(0.001)	(0.011)
Profes_Dev	-0.004**	-0.054	0.001	-0.068*
	(0.001)	(0.031)	(0.002)	(0.035)
Dimension	-0.004***	0.001	-0.001	0.004
	(0.001)	(0.015)	(0.001)	(0.017)
Ln_revenue	0.001	-0.611*	-0.001	-0.572
	(0.001)	(0.239)	(0.001)	(0.329)
Constant		0.739**		0.718*
		(0.239)		(0.320)
N	11580		9175	

Notes:
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Temporary variable omitted for collinearity in contract (base for *interim; apprenticeship; permanent; Member*)
Apprentice variable omitted for collinearity in position (base for *blue_collar; WhiteCollar; MiddleManager; Manager*)
Areas are handled as categorical complementary variables in the Oaxaca decomposition.

a. Multivariate analysis for the sample of workers where we observe education

In this sub-section, previous analyses are repeated for the subsample of workers with information on educational level. Table 6 repeats the analysis shown in table 5 but with additional education variables. The first relevant consideration is that, although the sample was cut by more than 70% (we still use only 3,125 observations of the 11,580 in Table 5) the adjusted R^2 remained essentially stable and suffered only a slight decrease. In fact, we still manage to explain almost 80% of the variations in hourly wages. Before delving into the interpretation of the coefficients in tables 6 and 7, it is noteworthy to mention that the results are not directly comparable with those in tables 4 and

5 as we are dealing with the special subsample with the educational level which could not be representative of the initial population (especially in terms of firms).

We observe that the level of education is an important variable in determining wages: higher levels of education, such as a bachelor's degree or postgraduate training, positively and significantly affect wages. On average, individuals with a university degree earn nearly 17%[71] more in hourly wages, but the returns to educational attainment are higher for men than for women and we obtain similar dynamics for those with postgraduate degrees. The lower effects on wages of a high degree for females could be due to several factors: limited women's opportunities for career advancement, training, or promotion, and undervaluing "female-typed" jobs which consist in a lower degree recognition and can also be related to the different fields of education. In addition, there is evidence to suggest that women are less likely to negotiate for higher wages and may not feel as entitled to higher pay as men.

Table 6 – Wage Equations by gender and hours of work

	All workers (At least 20h)			**At least 30h**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	All	**M**	**F**	**All**	**M**	**F**
Female	-0.095***			-0.092***		
	(0.009)			(0.009)		
Age	0.039***	0.039***	0.045***	0.039***	0.039***	0.045***
	(0.004)	(0.004)	(0.006)	(0.003)	(0.004)	(0.006)
Agesq	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Seniority	0.009***	0.008***	0.010***	0.009***	0.008***	0.012***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Senioritysq	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Middleschool	-0.012	-0.006	-0.008	-0.014	-0.007	-0.008
	(0.034)	(0.048)	(0.049)	(0.034)	(0.048)	(0.048)
Lowhighschool	-0.052	-0.026	-0.026	-0.050	-0.028	-0.031
	(0.035)	(0.047)	(0.057)	(0.035)	(0.046)	(0.057)
Highschool	0.014	0.038	-0.007	0.016	0.040	-0.007
	(0.032)	(0.044)	(0.046)	(0.032)	(0.044)	(0.045)
University	0.166***	0.182***	0.156**	0.169***	0.186***	0.158**
	(0.034)	(0.046)	(0.048)	(0.033)	(0.046)	(0.048)
Postgrad	0.156***	0.198***	0.140*	0.167***	0.201***	0.148*
	(0.040)	(0.055)	(0.057)	(0.040)	(0.054)	(0.057)
Interim	-0.078	-0.097	-0.076	-0.080	-0.094	-0.074
	(0.042)	(0.052)	(0.071)	(0.042)	(0.052)	(0.071)
Apprenticeship	-0.014	-0.025	-0.080	-0.015	-0.021	-0.080
	(0.050)	(0.058)	(0.103)	(0.050)	(0.057)	(0.102)
Permanent	0.030	-0.002	0.050*	0.024	0.001	0.040
	(0.016)	(0.021)	(0.025)	(0.016)	(0.021)	(0.025)

[71] 16.6 if they work at least 20h and 16.9 if they work at least 30h.

Member	-0.066	-0.123	0.056	-0.076	-0.125	0.046
	(0.055)	(0.073)	(0.100)	(0.054)	(0.072)	(0.099)
Blue_collar	0.116*	0.132	0.092	0.114*	0.131	0.096
	(0.047)	(0.074)	(0.066)	(0.047)	(0.073)	(0.066)
WhiteCollar	0.322***	0.341***	0.293***	0.317***	0.336***	0.290***
	(0.046)	(0.073)	(0.061)	(0.046)	(0.072)	(0.060)
MiddleManager	0.768***	0.785***	0.767***	0.765***	0.780***	0.762***
	(0.049)	(0.075)	(0.068)	(0.048)	(0.075)	(0.067)
Manager	1.422***	1.459***	1.399***	1.419***	1.454***	1.392***
	(0.051)	(0.077)	(0.074)	(0.050)	(0.077)	(0.074)
Smart	0.059***	0.047**	0.034	0.055***	0.046**	0.031
	(0.013)	(0.017)	(0.023)	(0.013)	(0.017)	(0.023)
Parttime	0.139***	0.232***	0.117***			
	(0.030)	(0.069)	(0.035)			
Purch_logistic	-0.012	-0.009	-0.008	-0.015	-0.012	-0.011
	(0.020)	(0.022)	(0.045)	(0.020)	(0.022)	(0.045)
Admin_fin	0.046*	0.012	0.073*	0.050**	0.005	0.084**
	(0.019)	(0.028)	(0.028)	(0.019)	(0.028)	(0.029)
Secretary_serv	0.081***	0.114***	0.060	0.101***	0.111***	0.091*
	(0.024)	(0.033)	(0.037)	(0.025)	(0.033)	(0.040)
Technical	0.100***	0.109***	0.085*	0.116***	0.113***	0.106**
	(0.020)	(0.025)	(0.040)	(0.020)	(0.025)	(0.041)
Commercial	0.020	0.009	0.046	0.021	0.009	0.050*
	(0.014)	(0.017)	(0.024)	(0.014)	(0.017)	(0.024)
Management	0.392***	0.553***	-0.503*	0.392***	0.549***	-0.490*
	(0.100)	(0.111)	(0.236)	(0.099)	(0.110)	(0.234)
Hr	0.175***	0.206***	0.162**	0.174***	0.204***	0.160**
	(0.038)	(0.053)	(0.054)	(0.037)	(0.052)	(0.053)
It	0.051	0.006	0.109	0.053	0.005	0.111
	(0.048)	(0.059)	(0.081)	(0.047)	(0.059)	(0.080)
Legal	0.042	0.008	0.090	0.071	0.011	0.153*
	(0.050)	(0.068)	(0.074)	(0.051)	(0.067)	(0.076)
Marketing	0.120***	0.139**	0.125**	0.122***	0.138**	0.127**
	(0.032)	(0.047)	(0.043)	(0.031)	(0.046)	(0.043)
Services	-0.170		-0.020	-0.166		-0.016
	(0.109)		(0.134)	(0.108)		(0.133)
Quality_safety	0.112*	0.100*	0.297	0.108*	0.095*	0.301
	(0.047)	(0.048)	(0.235)	(0.046)	(0.047)	(0.233)
Red	0.007	-0.056	0.084*	-0.004	-0.062	0.065
	(0.028)	(0.039)	(0.041)	(0.028)	(0.039)	(0.042)
High_FEM	-0.597***	-0.553***	-0.412	-0.592***	-0.551***	-0.420
	(0.071)	(0.079)	(0.219)	(0.070)	(0.078)	(0.217)
Constant	-11.199	-10.098	-8.400	-13.976	-12.574	-6.531
	(6.817)	(7.480)	(24.752)	(8.453)	(9.314)	(19.520)
N	3125	1995	1130	3068	1985	1083
Adjusted R^2	0.783	0.795	0.769	0.787	0.798	0.775

Notes:

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Primary variable omitted for collinearity in education (base for *Middleschool; Lowhighschool; Highschool; University; Postgrad)*

Temporary variable omitted for collinearity in contract (base for *interim; apprenticeship; permanent; Member*)

Apprentice variable omitted for collinearity in position (base for *blue_collar; WhiteCollar; MiddleManager; Manager*)
Primary variable omitted for collinearity in education (base for *Middleschool; Lowhighschool; Highschool; University; Postgrad*)
Operation variable omitted for collinearity in areas (base for *Purchase_logistic Administration_finance Technical Commercial Management HR IT Legal Marketing Secretary_services Services Quality_safety ReD*)

Table 7 should be interpreted in the same way as Table 5: the only things that change are the subgroups and the econometric process performed for getting the results displayed in columns 2 and 4.

Table 7 – Oaxaca Decomposition

	Al workers (At least 20h)				**At least 30h**			
	(1) Without correction		**(2) Heckman correction**		**(3) Without correction**		**(4) Heckman correction**	
PANEL A – Oaxaca decomposition into explained and unexplained part								
Males	2.932***		2.932***		2.931***		2.931***	
	(0.011)		(0.011)		(0.011)		(0.011)	
Females	2.857***		2.751***		2.854***		2.721***	
	(0.013)		(0.043)		(0.014)		(0.043)	
Difference	0.075***		0.181***		0.077***		0.210***	
	(0.017)		(0.044)		(0.017)		(0.044)	
	TOTAL		**TOTAL**		**TOTAL**		**TOTAL**	
	EXPL. (charact.)	**UNEXPL.**	**EXPL. - ADJUST** (charact.)	**UNEXPL. - ADJUST**	**EXPL.** (charact.)	**UNEXPL.**	**EXPL. - ADJUST** (charact.)	**UNEXPL. - ADJUST**
	-0.021	**0.095*****	**-0.021**	**0.202*****	**-0.015**	**0.092*****	**-0.015**	**0.225*****
	(0.016)	**(0.010)**	**(0.016)**	**(0.042)**	**(0.016)**	**(0.010)**	**(0.016)**	**(0.041)**
PANEL B - Breakdown of explained and unexplained parts into variables' contribution								
	EXPL. (character)	UNEXPL.	EXPL. - ADJUSTED (charact.)	UNEXPL. - ADJUSTED	EXPL. (character)	UNEXPL.	EXPL. - ADJUSTED (charact.)	UNEXPL. - ADJUSTED
Age	0.054***	-0.272	0.051**	-1.104*	0.065***	-0.250	0.063***	-1.282**
	(0.016)	(0.311)	(0.016)	(0.492)	(0.017)	(0.308)	(0.016)	(0.490)
Agesq	-0.047***	0.199	-0.044**	0.619*	-0.055***	0.195	-0.053***	0.714**
	(0.014)	(0.158)	(0.013)	(0.248)	(0.014)	(0.156)	(0.014)	(0.247)
Seniority	0.024***	-0.022	0.023***	-0.019	0.029***	-0.048	0.028***	-0.046
	(0.006)	(0.039)	(0.006)	(0.039)	(0.006)	(0.038)	(0.006)	(0.038)
Senioritysq	-0.004	-0.000	-0.003	-0.001	-0.006	0.014	-0.006	0.013
	(0.003)	(0.020)	(0.003)	(0.020)	(0.004)	(0.020)	(0.004)	(0.020)
Middleschool	0.000	-0.000	0.001	-0.005	0.001	-0.000	0.001	-0.006
	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)
Lowhighschool	-0.005	0.002	-0.007*	-0.008	-0.005	0.002	-0.007*	-0.008
	(0.003)	(0.007)	(0.003)	(0.009)	(0.003)	(0.007)	(0.003)	(0.008)
Highschool	0.001	0.017	-0.001	-0.026	0.001	0.017	-0.000	-0.037
	(0.002)	(0.018)	(0.003)	(0.025)	(0.002)	(0.018)	(0.003)	(0.025)
University	-0.023***	0.009	-0.020***	-0.073*	-0.026***	0.009	-0.022***	-0.094**
	(0.005)	(0.019)	(0.005)	(0.035)	(0.005)	(0.019)	(0.005)	(0.036)
Postgrad	-0.000	0.005	-0.000	-0.015	-0.000	0.004	-0.000	-0.021*
	(0.002)	(0.006)	(0.001)	(0.010)	(0.002)	(0.006)	(0.001)	(0.010)
Interim	-0.000	-0.000	-0.000	-0.000	0.000	-0.000	0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Apprenticeship	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Permanent	0.001	-0.044	0.001	-0.043	0.001	-0.033	0.001	-0.032
	(0.001)	(0.024)	(0.001)	(0.024)	(0.001)	(0.024)	(0.001)	(0.024)
Member	-0.001	-0.002	-0.001	-0.002	-0.001	-0.002	-0.001	-0.002

	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Blue_collar	0.016***	0.012	0.016***	0.014	0.016**	0.010	0.016**	0.014
	(0.005)	(0.020)	(0.005)	(0.020)	(0.005)	(0.020)	(0.005)	(0.020)
WhiteCollar	-0.056***	0.029	-0.057***	0.035	-0.055***	0.027	-0.055***	0.035
	(0.008)	(0.036)	(0.008)	(0.035)	(0.008)	(0.036)	(0.008)	(0.035)
MiddleManager	0.023**	0.002	0.023**	0.002	0.022**	0.002	0.022**	0.002
	(0.008)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)	(0.008)	(0.006)
Manager	0.024*	0.002	0.024*	0.003	0.022*	0.003	0.022*	0.003
	(0.010)	(0.004)	(0.010)	(0.004)	(0.010)	(0.004)	(0.010)	(0.004)
Smart	-0.013***	0.010	-0.013***	0.010	-0.012***	0.010	-0.012***	0.010
	(0.003)	(0.012)	(0.003)	(0.013)	(0.003)	(0.012)	(0.003)	(0.012)
PartTime	-0.005**	0.001	-0.005**	0.001				
	(0.002)	(0.002)	(0.002)	(0.002)				
High_FEM	0.026***	-0.011	0.027***	-0.012	0.028***	-0.011	0.028***	-0.012
	(0.007)	(0.008)	(0.007)	(0.009)	(0.007)	(0.009)	(0.007)	(0.009)
AREA' FE	yes	yes	yes	yes	yes	yes	yes	yes
FIRM' FE	yes	yes	yes	yes	yes	yes	yes	yes
Constant		-1.656		-3.064		-6.012		-7.222
		(14.465)		(14.476)		(15.496)		(15.494)
N	**3125**				**3068**			

Notes:
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Temporary variable omitted for collinearity in contract (base for *interim; apprenticeship; permanent; Member*)
Apprentice variable omitted for collinearity in position (base for *blue_collar; WhiteCollar; MiddleManager; Manager*)
Primary variable omitted for collinearity in education (base for *Middleschool; Lowhighschool; Highschool; University; Postgrad*)
Areas are handled as categorical complementary variables in the Oaxaca decomposition.

In fact, to obtain the results shown in the "even" columns, the Oaxaca decomposition (equation 2 in section 4 - methodology) was supplemented with Mill's ratio calculated according to the procedure in equation 3 in section 4 of the methodology. This procedure is important because it allows to correct female non-random selection in the labour market.

The first interesting result is that if we consider the level of education the part of the coefficient relating to characteristics (explained) turns in favour of females (although it is not significant). The breakdown of the explained part (panel B) reveals that -despite seniority and high positions remaining characteristics in favour of males- females are more educated in both the >20h and >30h sub-samples (significative coefficient of the university in the decomposition of the explained part).

The unexplained portion of the gap (potentially attributable to discrimination) remains higher and significative and in favour of men, moreover, consistently with the literature based on other sources of data, it rises considerably when Heckman's correction is applied (a sign that we were underestimating the magnitude and that indeed there is self-selection in female labour supply).

It is interesting to point out that – despite the higher values of the unadjusted explained part – the breakdown into variables reveals that virtually no variables have significant coefficients of significance: this could imply pure discrimination related solely to gender, regardless of the person's characteristics or position in the company, or it could also imply that there is a large part of the difference between the two groups that cannot be explained by the observable characteristics included in the regression model. This could be due to unobservable differences between the two groups that we are not able to account for because they are not in the model such as attitudes, efforts or social networks.

6. Conclusions

Taking into account the lower and self-selected female participation to the labour market, the literature shows a high and persistent gender wage gap at the disadvantage of women in Italy a gap that is much lower when one takes into account only the unadjusted wage gap.

By computing the unadjusted wage gap on average in the firms data used for our analysis, women earn only 78 per cent of male wages. This measure refers to the rough wage gap, which is statistically significant but it is based solely on gender information which is only able to explain around 6.7% of the variation in the hourly wage.

Preliminary analyses have found a negative effect of reduced working hours on hourly wages, which can be considered a form of indirect discrimination since this condition particularly affects women in our sample (37% against 5% of males). This trend is consistent with the negative and highly significative ($P < 0.01$) effect of part-time on hourly wages which are found in our regression analysis and in line with previous research (Schrenker 2023; Antonie *et al.* 2020).

For a better understanding of the wage gap, we introduce a regression analysis and the Oaxaca decomposition which allows us to control for many others dimensions and decompose the observed differential in explained and unexplained gap.

The regression analysis on the logarithm of hourly wage confirms the presence of a gender wage gap in favour of males which is in line with previous research on the Italian context based on other sources of data even if they do not employ firm data (Matteazzi & Scherer, 2021; Alfano *et al.* 2021, Furno, 2020; Piazzalunga & Di Tommaso, 2019).

In addition, other common trends resurface from our analyses such as age and seniority having a positive impact on wages (together with education introduced in subsection 6b).

In agreement with Cirillo and Ricci 2022, having a permanent contract has a significant and positive influence on hourly wages, especially for females and ,as expected, hourly salary increases with the increase of working positions (from apprentice, towards managerial positions). Moreover, as found by Centra and Cutillo (2009), both regression models and the Oaxaca decomposition confirm that being in a highly feminized occupation not only leads to lower general salaries but even lowers women’s returns.

As illustrated throughout the essay, a contribution to existing literature could be provided by the type and structure of the available data having detailed information on the firms' policies which allows us to analyse their impact of them on the gender wage gap.

So, it is worth noting to focus on firm fixed effects and their impact on wages estimated through the regression as on the discrimination component identified through the Oaxaca decomposition.

In this regard, higher wages are found to occur both for females and males when a higher share of women in managerial positions is observed. On the other hand, the same cannot be said for the gender composition of the board but, as discussed at length in Section 6 with extensive supporting literature, this dynamic may be linked to other factors related to both the gender composition of boards and the gender pay gap in Italy. Moreover, the Oaxaca decomposition shows that having a higher share of women on the Board is a firm's observable characteristic which is significatively in favour of women's wages.

Turning to regulatory policies aimed to reach a better work-life the main results are the following:

- Having formalized policies and services specifically for parenting is associated with higher overall wages. One possible explanation is that companies that offer these benefits are more likely to be larger, more established, and higher-paying firms that can afford to provide these services to their employees. These firms may have a better reputation, attract more talented workers, and have a more competitive job market, which can lead to higher wages for their employees. However, when analysing the Oaxaca decomposition by gender, it appears that this negatively affects females' wages but as discussed this should not necessarily be interpreted negatively: maybe because they can retain mothers (characterized by lower workload) and have lower self-selection of female workers.
- Having formalized flexibility policies is both a firm's characteristic in favour of females' wages and with higher returns for women, especially in the subsample which includes part time workers.
- The regression results and Oaxaca decomposition show that firms' part-time regulation has no significant effect on total wages within the firm. This suggests that any negative impact of part-time regulation is likely captured by the individual variable indicating whether a worker chooses to work part-time, rather than by the fixed effect of the firm itself. While this finding is encouraging, it also highlights the importance of considering individual-level factors when examining the effects of employment policies on wages and other labour outcomes.

Finally, the firm's dimension, revenues have a positive impact on women's salaries according to the Oaxaca decomposition together with having formalized policies on professional development. Evidence from the Heckman application, consistently with the literature, reveals that when correcting for the non-random self-selection of women in the labour market, the unexplained part of the wage gap (potentially attributable to discrimination) in favour of men is larger than when we do not correct for the non-random selection of women into the labour market (Addabbo, 2018; McKay & Mussida, 2018; Olivetti and Petrongolo, 2008).

Based on the evidence presented, there are several policy recommendations that could be implemented at both the firm and national levels to address the gender gap (in terms of wages but not only).

At the firm level, companies could implement policies to increase the share of women in managerial positions, as this has been shown to have a positive impact on both female and male wages. Additionally, formalized policies and services specifically for parenting and flexible work arrangements could also be implemented to support working parents, particularly women. It is important to carefully consider the potential impact of these policies on female wages, as the evidence suggests that while these policies may be associated with higher overall wages, their impact on female wages may be misunderstood as negative due to mothers retention and the lower self-selection of female workers. Firms could also focus on professional development opportunities, as evidence suggests that having formalized policies on professional development has a positive impact on women's salaries.

At the national level, policies could be implemented to support working parents, such as providing subsidies for childcare and parental leave. Additionally, policies could be implemented to promote gender diversity in managerial positions and on boards of directors. This could include setting quotas or targets for female representation, as has been done in Italy for some kind of companies. Finally, policies could be implemented to address discrimination in the labour market, including strengthening anti-discrimination laws and providing training for managers and recruiters on unconscious bias.

If the nature of the database allows us to have interesting insights, on the other hand, the database used limits the extension of our results for three main causes.

At first, the sample used is rather limited in size (12 firms and 16 legal entities) and it is not randomly selected. The sample is indeed not representative of the Italian firms' population as it consists of firms of bigger dimensions than the average firms' dimension observable in Italy and it includes firms that are located in the northern part of Italy and refer also to a limited set of NACE. Moreover, are firms which are self-selected in a process of gender equality evaluation: all factors that most likely lead to an underestimation of the gender gap.

This non-random selection is also reflected in the composition of workers deviating slightly from the national average (lower share of part-time workers and higher share of permanent contracts).

Finally, some variables, that are shown to impact of wages in the literature, are observed only for part of the sample (like education) or proxied (for instance not observing the actual number of children we proxy this variable with the number of dependents).

This set of limitations related to the available data implies that caution should be exercised in generalizing the results beyond the same context.

We plan to extend the analysis by collecting more firms' observations, acquiring more information on the workers' levels of education and estimating quantile regressions to analyse the path of wage discrimination over the wage distribution and its relationship with the observed evidence of glass ceiling and sticky floor effects.

7. References

Addabbo, T., and Favaro, D. (2011) ‘Gender wage differentials by education in Italy’. *Applied Economics*, 43(29), 4589–4605.

Addabbo, T. (2018) ‘Disuguaglianze di genere nel lavoro e nelle retribuzioni in Italia’, Ch.8 in Maurizio Franzini and Michele Raitano (eds) *Il mercato rende diseguali? La disuguaglianza dei redditi in Italia*, Bologna, Il Mulino, pp.173-192.

Addison, J.T., Ozturk, O.D. and Wang, S. (2018) ‘The Occupational Feminization of Wages’, *ILR Review*, 71(1), pp. 208–241.

Alfano, V., Cicatiello, L. , Lucio Gaeta g., and Pinto M..)(2021). ‘The Gender Wage Gap among Ph.D. Holders: Evidence from Italy’. *The B.E. Journal of Economic Analysis & Policy* 21 (3): 1107–48. https://doi.org/10.1515/bejeap-2020-0319.

Antonie, Luiza, Laura Gatto, and Miana Plesca. (2020). ‘Full-Time and Part-Time Work and the Gender Wage Gap’. *Atlantic Economic Journal* 48 (3): 313–26. https://doi.org/10.1007/s11293-020-09677-z.

Astorne-Figari, C., and Speer Jamin D. (2019). ‘Are Changes of Major Major Changes? The Roles of Grades, Gender, and Preferences in College Major Switching’. *Economics of Education Review* 70 (June): 75–93. https://doi.org/10.1016/j.econedurev.2019.03.005.

Becker, Gary S. (1971). The Economics of Discrimination. Edited by 2d edition. Economic Research Studies. Chicago, IL: University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/E/bo22415931.html.

Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2019). ‘Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway’. *Review of Economic Studies*, 86(1), 191–239.

Blau D.F. and Kahn, L.M. (2017) ‘The Gender Wage Gap: Extent, Trends, and Explanations’, *Journal of Economic Literature*, 55(3), 789–865

Booth, Alison L., Cardona-Sosa L., and Nolen P. (2018). ‘Do Single-Sex Classes Affect Academic Achievement? An Experiment in a Coeducational University’. *Journal of Public Economics* 168 (December): 109–26. https://doi.org/10.1016/j.jpubeco.2018.08.016.

Busch, F. (2020) ‘Gender Segregation, Occupational Sorting, and Growth of Wage Disparities Between Women’, Demography (Springer Nature), 57(3), pp. 1063–1088.

Busch, A. and Holst, E., (2011) ‘Gender-Specific Occupational Segregation, Glass Ceiling Effects, and Earnings in Managerial Positions: Results of a Fixed Effects Model’, IZA Discussion Paper No. 5448.

Campa, D., & Quaglione, D. (2019). ‘Women on boards and financial performance in Italian public companies.’ *Journal of Management and Governance*, 23(4), 933-957. doi: 10.1007/s10997-018-9422-2.

Card, D., Cardoso, A.R., Kline, P. (2016) 'Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women.', *The Quarterly Journal of Economics* 131(2):633–686

Carlana, M. , and Corno L. (2021). 'Parents and Peers: Gender Stereotypes in the Field of Study'. 16582. *CEPR Discussion Papers*. CEPR Discussion Papers. C.E.P.R. Discussion Papers. https://ideas.repec.org/p/cpr/ceprdp/16582.html.

Carlana, M. (2019). 'Implicit Stereotypes: Evidence from Teachers' Gender Bias'. *The Quarterly Journal of Economics* 134 (3): 1163–1224. https://doi.org/10.1093/qje/qjz008.

Centra, M. and Cutillo, A. (2009) Differenziale salariale di genere e lavori tipicamente femminili, *Studi Isfol*, no. 2.

Cirillo, V., and Ricci A. (2022). 'Heterogeneity Matters: Temporary Employment, Productivity and Wages in Italian Firms'. Economia Politica 39 (2): 567–93. https://doi.org/10.1007/s40888-020-00197-2.

Del Giudice, M., Scuotto, V., & Carayannis, E. G. (2020). 'A gender diversity index for board of directors in the Italian context.' Journal of Open Innovation: Technology, Market, and Complexity, 6(2), 34. doi: 10.3390/joitmc6020034.

European Commission. (2021). Gender pay gap statistics. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics

Fitzenberger, B., and Kunze, A. (2005) 'Vocational Training and Gender: Wages and Occupational Mobility among Young Workers', *Oxford Review of Economic Policy*, 21 (3), 392-415.

Fort, M., & Schneeweis, N. (2013). Decomposing gender wage gaps in a developing country: Evidence from the Philippines. Journal of Labor Economics, 31(2), 381-414. doi: 10.1086/669158.

Furno, M. 2020. 'Returns to Education and Gender Wage Gap Across Quantiles in Italy'. *Central European Journal of Economic Modelling and Econometrics* 12 (2): 145–69.

Gregorič, A., & Cernoša, A. (2019). 'Women on boards and company performance: Evidence from Italy.' *Gender in Management: An International Journal*, 34(6), 492-510. doi: 10.1108/GM-01-2018-0012.

Goldin, C., Kerr, S., Olivetti, C., Barth, E. (2017) 'The expanding gender earnings gap: evidence from the LEHD-2000 census'. *American Economic Review Papers and Proceedings* 107(5):110–114

Heckman, J. J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Econometrics and Social Measurement 5: 475–492.

Heckman, J. J. (1979) Sample selection bias as a specification error. Econometrica 47: 153–161.

International Labour Organization (ILO). (2019). Women and men in the informal economy: A statistical picture (3rd ed.). Geneva: International Labour Office. Retrieved from https://www.ilo.org/global/publications/books/WCMS_626831/lang--en/index.htm

Istituto Nazionale di Statistica (ISTAT). (2021a). Lavoro e retribuzioni. Indagine sulle forze di lavoro [Work and earnings. Labor force survey]. https://www.istat.it/it/files/2021/08/Lavoro-e-retribuzioni-II-trimestre-2021.pdf

Istituto Nazionale di Statistica (ISTAT). (2021b). Rapporto sulle imprese 2021. Edited by Andrea de Panizza. 1st ed. IT: Istat. https://doi.org/10.1481/Istat.Rapportoimprese.2021.

Jann, Ben. (2008). 'The Blinder–Oaxaca Decomposition for Linear Regression Models'. The Stata Journal: Promoting Communications on Statistics and Stata 8 (4): 453–79.

Lavy, V. (2008). 'Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment'. *Journal of Public Economics*, 92 (10-11): 2083-2105. doi:10.1016/j.jpubeco.2008.02.009

McKay, S., and Mussida, C. (2018). 'Wages And Employment Chances For Women: The Role Of Expected And Actual Fertility'. *Rivista Internazionale Di Scienze Sociali* 126 (3): 255–82.

Matteazzi, E., and Stefani Scherer. 2021. 'Gender Wage Gap and the Involvement of Partners in Household Work'. *Work, Employment and Society* 35 (3): 490–508. https://doi.org/10.1177/0950017020937936.

Mincer, Jacob. (1974). 'Schooling, Experience, and Earnings.' *Human Behavior & Social Institutions No. 2*. National Bureau of Economic Research, Inc.

Mussida, C., and Sciulli, D. (2023). 'Low Pay and Household Poverty in Italy'. *Working Papers 635*, ECINEQ, Society for the Study of Economic Inequality, February. https://ideas.repec.org//p/inq/inqwps/ecineq2023-635.html.

Mussida, C., and Picchio, M. (2014) 'The gender wage gap by education in Italy.' *Journal Of Economic Inequality*, 12(1), 117-147.

Oaxaca, R. (1973). 'Male-Female Wage Differentials in Urban Labour Markets', International Economic Review 14: 693–709.

Olivetti, C. and Petrongolo, B. (2008) 'Unequal pay or unequal employment? A cross-country analysis of gender gaps', *Journal of Labour Economics*, 26 (4), 621-654.

Pastore, F., & Verashchagina, A. (2015). 'Gender wage differentials and occupational distribution in Italy: A review of the literature'. Journal of Labor Research, 36(3), 235-248. doi: 10.1007/s12122-015-9202-2.

Piazzalunga D.; Di Tommaso, M.L., (2019). "The increase of the gender wage gap in Italy during the 2008-2012 economic crisis," The Journal of Economic Inequality, Springer;Society for the Study of Economic Inequality, vol. 17(2), pages 171-193, June.

Schrenker, Annekatrin. (2023). 'Do Women Expect Wage Cuts for Part-Time Work?' Labour Economics 80 (January): 102291. https://doi.org/10.1016/j.labeco.2022.102291.

Verbeek, Marno. (2014) 'A Guide to Modern Econometrics'. *Applied Econometrics* 8 (February): 125–32.

Zizza R., (2013) "The gender wage gap in Italy", *Questioni di Economia e Finanza* (Occasional Papers) 172, Bank of Italy.

8. Appendix

Appendix A1 – Sample characteristics and nature of self – selection

As anticipated in the introduction, even if we are dealing with a big and exhaustive database at the employee level, we must raise awareness that it is not a firm level.

As anticipated these are companies that show interest in gender equality and assume they can get certified, and this kind of self-selection could lead us to underestimate the phenomenon.

Moreover, aside from this fact, the dimension of the enterprises is relatively big (Mainly big enterprises, only some of the middle size, a small one[72] and no one in the micro category) if compared to the national average in which more than 60% of enterprises have only one employee (usually sole proprietorships with the owner self-employed), 30% are microenterprises, small enterprises are about 5%, and medium and large enterprises together are less than 0.7% (ISTAT 2021b).

Moreover, firms are representative only of 13 ATECO codes (NACE) out of the 99 present and some sections aren't represented (eg: (a) agriculture, forestry and fisheries ; (b) extraction of minerals from quarries and mines; (f) construction; etc.) but at least they covered heterogeneous physiologic characteristics (share of females, required tasks, duties, products, sectors). In addition, all the companies are headquartered in northern Italy, which we know has different characteristics than the rest of the peninsula, especially in terms of female labour force participation.
Finally, even if the employee population is s exhaustive within the analysed firm, has some insight into the descriptives, not in line with the national averages and most likely due to the characteristics of enterprises (location, size, etc.). The proportion of permanent contracts among the analysed workforce is higher than expected (94%), with an even greater share observed among female employees. This finding could lead to an underestimation of the real gender wage gap. It is worth noting that the national average for permanent contracts in 2022[73] is 83%, with a slightly lower rate among females (81%) than males (84%). The descriptive statistics show also that only 5% of males have a part-time Job against the 37% of females with a gap of 32%). A high share of women in part-time positions is expected as this phenomenon in Italy is one of the highest in Europe (ILO 2019) and according to Istat (2021a) the share of females working part-time in Italy was 47.5%, compared to 16.2% of males and these percentages have remained relatively stable over the past

[72] According to the size definition of the Italian Decree of the April 18, 2005 Adjustment of the criteria for identifying small and medium-sized enterprises to the Community Framework. (GU Serie Generale n.238 del 12-10-2005). https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=2005-10-12&atto.codiceRedazionale=05A09671&elenco30giorni=false.

[73] http://dati.istat.it/Index.aspx?QueryId=26887# .

few years. Although there are fewer part-time workers in our database the gender proportions are

	(1) FEMALE		(2) MALE		(3) T-TEST ($\bar{x}_m - \bar{x}_f$)	
	mean	**sd**	**mean**	**sd**	**b**	**t**

consistent with national averages (a 30% higher gap in favour of women)

Income	40121.26	25153.09	44607.61	32133.47	4486.35***	(5.56)
Yhour	19.83	12.28	21.58	15.68	1.75***	(4.43)
Age	42.39	10.08	44.76	10.29	2.37***	(8.05)
Seniority	10.13	9.67	13.69	10.60	3.56***	(12.26)
Primary	0.02	0.15	0.02	0.12	-0.01	(-1.67)
Middleschool	0.07	0.25	0.03	0.18	-0.04***	(-4.13)
Lowhighschool	0.09	0.28	0.18	0.38	0.09***	(7.65)
Highschool	0.32	0.47	0.41	0.49	0.09***	(5.33)
University	0.42	0.49	0.28	0.45	-0.14***	(-7.89)
Postgrad	0.09	0.28	0.08	0.28	-0.00	(-0.21)
Interim	0.01	0.08	0.01	0.08	0.00	(0.05)
Apprenticeship	0.01	0.10	0.01	0.09	-0.00	(-0.23)
Temporaryfull	0.13	0.34	0.10	0.30	-0.03**	(-3.22)
Permanent	0.86	0.35	0.88	0.32	0.03**	(2.65)
Member	0.00	0.07	0.01	0.09	0.00*	(2.11)
Apprentice	0.01	0.10	0.00	0.05	-0.01*	(-2.55)
Blue_collar	0.18	0.38	0.31	0.46	0.13***	(10.58)
White-collar	0.72	0.45	0.57	0.50	-0.15***	(-11.48)
Middle-man.	0.06	0.25	0.09	0.28	0.02**	(2.84)
Manager	0.02	0.15	0.04	0.18	0.01**	(2.69)
Purch_logistic	0.02	0.14	0.04	0.20	0.02***	(4.62)
Admin_fin	0.15	0.36	0.06	0.23	-0.09***	(-9.98)
Secretary_serv	0.03	0.18	0.02	0.13	-0.01**	(-2.99)
Technical	0.03	0.17	0.08	0.28	0.05***	(8.40)
Commercial	0.34	0.47	0.32	0.47	-0.01	(-1.04)
Management	0.00	0.05	0.00	0.04	-0.00	(-0.51)
Hr	0.01	0.12	0.01	0.08	-0.01**	(-2.65)
It	0.01	0.09	0.01	0.11	0.00	(1.71)
Legal	0.01	0.10	0.00	0.06	-0.01*	(-2.13)
Marketing	0.03	0.17	0.01	0.10	-0.02***	(-4.58)
Services	0.03	0.16	0.01	0.09	-0.02***	(-4.45)
Quality_safety	0.00	0.03	0.01	0.08	0.01***	(3.59)
Red	0.08	0.27	0.10	0.30	0.02*	(2.10)
Operations	0.26	0.44	0.33	0.47	0.07***	(5.38)
Smart	0.55	0.50	0.30	0.46	-0.25***	(-17.46)
Parttime	0.04	0.19	0.01	0.08	-0.03***	(-6.77)
High_fem	0.17	0.38	0.06	0.24	-0.11***	(-11.30)
Hweek	39.42	4.05	40.29	2.18	0.87***	(8.55)
N	1858		3320		5178	

Appendix A2 – Gender comparison on personal characteristics and firm's allocation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix A3 – Horizontal and vertical segregation, details by gender

Area	Income / ln(Your) / freq.											
	Male						Female					
	apprentice	blue_collar	impiegato	quadro	dirigente	Total	apprentice	blue_collar	impiegato	quadro	dirigente	Total
Purc_log		27728.63	40909.64	69352.87	233891.3	**47937.05**		27601.78	33417.4	57668.18	109331	**36620.74**
		2.571	2.919	3.479	4.587	**2.918**		2.565	2.775	3.28	3.97	**2.832**
		68	71	18	8	**165**		4	64	8	1	**77**
Admin_fin		32309.83	39454.11	72189.56	149115.3	**56648.17**		26718.24	35831.74	70014.85	154937.5	**43489.22**
		2.746	2.922	3.532	4.209	**3.155**		2.562	2.886	3.507	4.236	**2.993**
		12	150	36	25	**223**		3	272	30	12	**317**
Technical		31738.25	41610.51	68510.92	146404.2	**44006.85**			36676.02	55794.25		**38353.06**
		2.705	2.927	3.444	4.144	**2.95**			2.867	3.297		**2.905**
		76	162	33	5	**276**			52	5		**57**
Commercial	24440.51	29655.31	40734.99	78998.81	191082.1	**48433.43**	23288.26		37830.33	70489.23	180748.1	**42039.95**
	2.434	2.668	2.948	3.627	4.423	**3.049**	2.383		2.881	3.514	4.364	**2.936**
	4	9	941	101	31	**1086**	7		594	34	12	**647**
Management		25191.04	67606.16		392505.3	**247203.5**			29182.78	62593.06	68456.96	**40710.47**
		2.502	3.45		5.223	**4.328**			2.688	3.379	3.502	**2.926**
		1	2		4	**7**			10	4	1	**15**
HR			39044.18	73755.83	195444.7	**67272.16**		20676.28	34638.18	69327.66	137253	**53490.05**
			2.884	3.56	4.42	**3.245**		2.438	2.808	3.485	4.169	**3.103**
			25	10	5	**40**		1	35	12	6	**54**
IT			47285.75	74467.48	142521.4	**64420.2**			44571.32	77481.16		**56186.56**
			3.104	3.562	4.195	**3.341**			3.056	3.613		**3.252**
			31	14	5	**50**			11	6		**17**
Legal			37661.75	67390.3	138611.4	**52287.55**			38352.89	80248.63		**48210.71**
			2.892	3.501	4.207	**3.134**			2.91	3.655		**3.085**
			9	3	1	**13**			13	4		**17**
Marketing		31397.45	42485.11	83133.71	168988.7	**70408.45**			39658.93	75740.38	153081.6	**56451.89**
		2.599	2.976	3.686	4.364	**3.391**			2.92	3.594	4.283	**3.177**
		2	16	18	3	**39**			46	15	5	**66**
Operations	23078.02	24918.99	36824.31	66556.16	160828.3	**27617.8**	23881.81	16802.21	25568.22	69812.94	173606.2	**18349.3**
	2.349	2.509	2.827	3.45	4.309	**2.568**	2.431	2.367	2.583	3.506	4.267	**2.402**
	4	3079	385	51	21	**3540**	2	3676	472	29	5	**4184**
Secret_serv		32952.22	43019.32	69862.98	154496	**41894.25**	25106.78	16155.32	36962.84	63200.52		**33077.05**
		2.734	2.998	3.516	4.316	**2.935**	2.419	2.436	2.888	3.84		**2.796**
		23	33	2	1	**59**	4	10	44	1		**59**
Services			35211.48	69707		**37590.48**			31899.58	72088.89		**34310.94**
			2.823	3.518		**2.871**			2.736	3.557		**2.785**
			27	2		**29**			47	3		**50**
Qual_safe		29776.45	50888.01	98200	176456	**43718.23**			55100			**55100**
		2.638	3.167	3.862	4.448	**2.885**			3.297			**3.297**
		16	5	1	1	**23**			2			**2**
ReD	22554.22	26364.75	39231.87	62684.86	94379.33	**46181.52**	24246.94	26576.28	38286.14	60873.51	73932.49	**39835.46**
	2.371	2.541	3.014	3.481	3.831	**3.126**	2.456	2.582	3.035	3.476	3.63	**3.054**
	2	10	234	55	20	**321**	4	1	131	10	2	**148**
Total	**23518.25**	**25262.08**	**39920.38**	**72049.33**	**167353.4**	36367.47	**24011.55**	**16823.89**	**33965.53**	**69282.39**	**154791.8**	24770.22
	2.387	**2.519**	**2.93**	**3.539**	**4.284**	2.764	**2.415**	**2.368**	**2.807**	**3.502**	**4.22**	2.552
	10	**3296**	**2091**	**344**	**130**	5871	**17**	**3695**	**1793**	**161**	**44**	5710