

"STEM in Genere": An Impact Evaluation

Abbiati Giovanni, Leggerini Chiara and Bannò Mariasole

University of Brescia, Brescia, Italy

giovanni.abbiati@unibs.it

chiara.leggerini@unibs.it

mariasole.banno@unibs.it

Abstract: Research on labour market disparities between men and women has long since identified horizontal segregation in the educational system as one of the main factors driving the gender pay gap. One of the factors identified in the literature believed to be at the heart of horizontal gender segregation is the lack of representation and visibility of women in STEM fields. In this paper, we investigate the effectiveness of a program named "STEM in Genere", developed within the Gender Equality Plan 2022-2024 of the University of Brescia (Italy), which is aimed at contrasting the underrepresentation and the stereotypical representation of women in science among primary and lower secondary school students. Students of participant classes will meet an educator for a total of 2 hours, in which counter-stereotypical thinking of women in science will be stimulated via learning games. Teachers of participant classes will follow in parallel a dedicated workshop about non-stereotypical science teaching. The effects of the program on students' gender stereotypes in science and study/job aspirations will be evaluated via a randomised controlled trial (RCT) conducted in primary and secondary schools of the Brescia province, an affluent province located in the northern part of Italy. In particular, we will randomise classes within participating schools to have a good balance between the statistical power of the experiment and potential contamination threats. The contribution of this evaluation is threefold. First, we will collect first-hand data on students' views and aspirations in a country - Italy - in which the population traditionally holds quite conservative views about gender roles. Second, we will provide robust, experimental evidence on the effectiveness of the program "STEM in Genere", which is potentially scalable nationwide. Lastly, we will contribute to the methodological debate about question framing in survey methodology, by randomly varying the formulation of the questions in the surveys. Policy implications for educational and learning environments will be discussed.

Keywords: Gender gap; STEM; RCT

1. Introduction

The educational landscape of modern societies is characterised by the presence of both vertical (i.e. related to attainment levels) and horizontal (related to types within levels) gender inequalities. While vertical gender inequalities have reversed over the last two decades, now favouring girls over boys (Lindberg et al., 2010), horizontal inequalities within secondary and tertiary education demonstrate remarkable persistence, with STEM-related fields show a low representation of females. In contrast, humanities exhibit a lower representation of males (OECD, 2021). Even among STEM fields, females tend to select those fields that are less math-intensive and leading to care professions, such as Medicine and STEM teaching (Barone, 2011).

Many commentators have emphasized that the humanistic-STEM and care-technical divides significantly contribute to the gender pay gap. Fields dominated by women often lead to less lucrative careers than those in technical fields, such as computing and engineering (Blau and Kahn, 2017). The challenges women face in the labor market also produce negative economic externalities. In Italy, where this study's case is situated, a recent report from the Bank of Italy estimated that the underrepresentation of females in the labor market, and the significant loss of human capital this entails incurs a significant cost in terms of GDP (Carta et al., 2023).

The factors contributing to horizontal segregation, particularly the gender gap in STEM fields, remain debated within the scientific community. The literature concurs that the underrepresentation of women in STEM is not due to inferior mathematical reasoning abilities (Lindberg et al., 2010), a notion commonly held in the past. In open and democratic societies, this phenomenon is typically attributed to differing interests between men and women. However, the root of these differences is contentious. Early researchers primarily focused on biological predispositions towards caregiving (Lueptow et al., 1995); see also the debate following, (Kahlenberg and Wrangham, 2010), but recent studies increasingly examine the impact of traditional gender stereotypes in shaping societal expectations for appropriate attitudes and behaviors in men and women (e.g. Guiso et al., 2008). This cultural influence is thought to either mask inherent motivations and interests as cultural norms (Cheryan, 2012) or to amplify and reinforce potential biological predispositions for caregiving between men and women (Hooven, 2021). In education, this phenomenon is exemplified by stereotypes that associate STEM fields with typically male characteristics (Diekman et al., 2010; Makarova et al., 2019) or by the endorsement of *gender essentialism* (i.e. the beliefs that man and women have inherently different inclinations and should behave accordingly) by the teaching workforce (Levanon and Grusky, 2016).

There is also a lack of consensus in the literature regarding the effectiveness of interventions to counteract stereotypical perceptions of gender in STEM careers. Various theoretical frameworks emphasize the importance of adopting non-stereotypical beliefs early in life to shape career and educational aspirations, which begin to form around ages 11 or 12 (Kang et al., 2019). Specifically, the Social Cognitive Career Theory (Lent et al., 2000), building on a's social cognitive theory (Bandura, 1986), suggests that self-efficacy beliefs and outcome expectations influence career choices. Similarly, Eccles' model of achievement-related choices posits that stereotypes indirectly influence educational and career choices through biased estimations of success among available options (Eccles, 2009). Social role theory (Eagly and Wood, 2012) proposes that stereotypical beliefs arise from observing patterns in social roles and occupations, with the prevalence of men in STEM fields reinforcing implicit associations that underpin sectoral preferences.

A common inference from these theories is that exposure to female STEM professionals might alter the association between gender and STEM preferences, either by enhancing young women's perceived likelihood of success in STEM careers or by disrupting the implicit link between maleness and STEM. In recent decades, the focus has been on presenting young students with successful female role models. Despite their popularity and frequent recommendation in educational practices, scant evidence supports their effectiveness. Furthermore, the wide variety of formats in which role model interventions are implemented complicates the identification of factors that enhance their effectiveness (De Gioannis et al., 2023).

In this study, we contribute to the research examining the effectiveness of role model interventions in mitigating the gender gap in STEM fields. Our focus is on "STEM in Genere," a training program aimed at influencing both teachers' and students' perceptions of women's roles in science. This intervention, targeting primary and secondary school students, is funded by the University of Brescia. It will be evaluated through a randomized controlled trial during the 2024-2025 school year, involving a sample of primary and lower secondary students in the Italian province of Brescia. Our analysis will center on the short-term effects of the intervention on gender stereotypes and educational aspirations. Additionally, we plan to introduce random variations in the wording of our tests. This methodological choice is intended to assess language's influence on eliciting gender stereotypes in questionnaires, specifically those written in Romance languages. Furthermore, we will discuss policy implications and provide cost-effectiveness estimates concerning the potential scaling-up of this project.

2. Institutional Context

The educational system in Italy consists of three main cycles: primary (grades 1 to 5), lower secondary (grades 6 to 8), and upper secondary (starting at 9th grade). The primary and lower secondary cycles do not implement tracking, ensuring a uniform curriculum for all students. Tracking begins at the onset of the upper secondary cycle, where families can select a specific track for their children. This phase of education is characterized by significant differentiation, both vertically, in terms of preparation for university, and horizontally, across the humanistic and STEM disciplines. Consequently, schools and classes tend to exhibit considerable homogeneity in both social and gender composition (Guetto and Vergolini, 2017; OECD, 2021). Within this educational framework, the province of Brescia in Northeastern Italy presents a typical socio-economic scenario: a robust industrial sector dominated by small and medium-sized family-run manufacturing enterprises (ISTAT, 2023), alongside a population adhering to relatively traditional values (Lomazzi, 2017). These values often reinforce traditional gender roles, both within family settings and the workplace.

3. Empirical Analysis

3.1 Case Study

The "STEM in Genere" project, hereafter referred to as 'the treatment', is an initiative under the University of Brescia's Gender Equality Plan (GEP) for 2022-2024. This project aims to counteract gender stereotypes in science education among primary and lower secondary students in Brescia. The focus on these educational levels stems from the need to address gender stereotypes in science before the onset of tracking, which is the main factors driving gender differences in enrolment tertiary education in Italy (Barone and Assirelli, 2020). The treatment consists of interactive sessions where trained tutors engage students with educational games about science, scientists' work, and women's role in scientific history. This approach challenges the notion of science as a male-exclusive field. An adapted version of the game "Guess Who?" highlights the lives and achievements of notable scientists of both genders. Also the narrative of science's life, through the instrumentation used and which characterizes specific scientific subject will be the means of conveying the message of gender inclusion in

Science. The opportunity to hear a story, accompanied by the attraction and fascination that illustration exerts on little boys and girls, will be the most effective way to capture the attention of young listeners while keeping them entertained. The duration of these sessions ranges from a minimum of 2 to a maximum of 4 hours per classroom, with materials customized to the students' age group. Additionally, teachers participate in a dedicated training course (Manzella et al., 2023), which has effectively reduced gender stereotypes in educational settings. This training aims to align the project's objectives with the teaching practices in participating classes.

3.2 Evaluation Design

We have developed a randomized controlled trial (RCT) with a delayed treatment design to evaluate the program's effectiveness on students' stereotypical views and aspirations. In the spring of 2024, we will recruit a sample of randomly selected primary and lower secondary schools in the Brescia province. Participating schools must ensure the involvement of at least four classes. In return, the research team will provide each school with a report containing aggregated data about their students' views and aspirations. It will deliver the program to all participating classes within the school year. Randomization will be performed at the classroom level: within each school half of the classrooms will be assigned to the treatment and half to the control group. This design balances the need for statistical power with potential contamination risks. We aim to minimize contamination by delaying the treatment for control classes by three to four months. Data collection will occur via two surveys: a baseline survey at the project's commencement and a follow-up survey approximately four months before administering the treatment to control classes. Figure 1 graphically represents the evaluation process, with horizontal lines tracing each group's journey and dotted vertical lines indicating survey administrations (pre- and post-treatment). The surveys will be conducted simultaneously for both groups, as shown by the red dots.

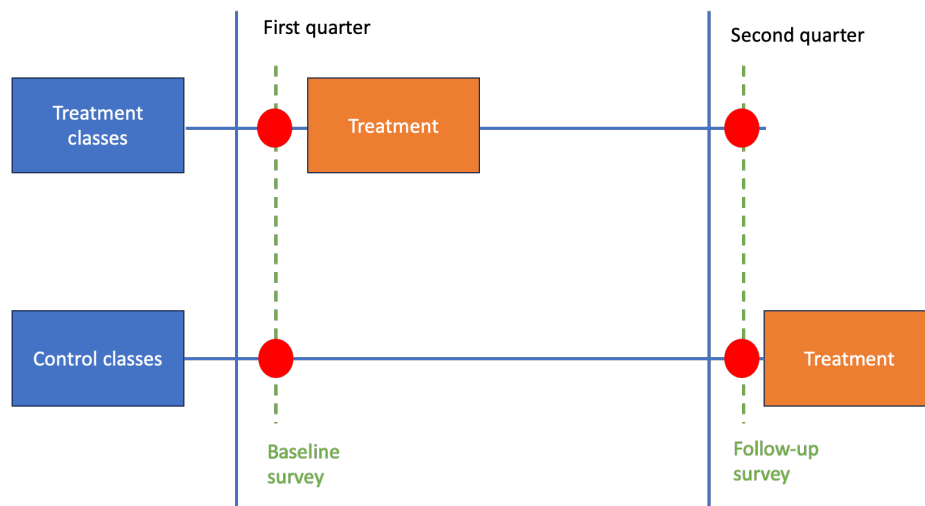


Figure 1: Evaluation design, source: personal elaboration.

The effects will be estimated thanks to the misalignment between the intervention's timing and the scheduling of questionnaire administration in the two groups. Impact estimates will be derived by regressing our outcome measures (see § 2.3) on the treatment status of the classroom, controlling for pre-treatment outcome values and school fixed effects. The treatment effects will be estimated using the following model:

$$(1) Y_{ijcs1} = \alpha + \beta T_{jcs} + \delta Y_{ijcs0} + \gamma S_s + \theta_j + \varepsilon_{ijcs}$$

Y_{ijcs1} represents the outcome of the i^{th} student in enrolled in class c of grade j of school s at the follow-up; T_{jcs} represents the treatment indicator; Y_{ijcs0} represent the outcome variable measured at baseline, S_s and θ_j represent school and grade fixed effects respectively, and ε is the error term. Standard errors are clustered at the school level.

3.3 Measures

A key aspect of this evaluation is the inclusion of classes from first to eighth grade, which vary widely in reading and cognitive abilities. The project aims to assess the program's impact on endorsing gender stereotypes related to science and scientists. These will be measured using explicit scales via questionnaires (for 4th to 8th graders)

and indirect tests for 1st to 3rd graders (for a review of instruments, see De Gioannis et al., 2023). Indirect tests, like a modified version of the Draw A Scientist Test (DAST) (Finson, 2002), will assess the main outcome among 1st to 3rd graders, an age in which reading a text and answering a survey might be too demanding. For 6th to 8th graders, questions about job aspirations will be included as a secondary outcome.

3.4 Methodological Experiment

Regarding our primary outcome, we will embed a survey experiment in our version of the “Draw A Scientist Test”. Issues related to wording in questionnaires have been extensively studied and debated (Schaeffer and Dykema, 2020). Wording of the questions is particularly consequential in our case, since in Romance languages the grammar masculine in its plural form is intended to serve as neutral, thus including also women. To assess the impact of using grammatically masculine versus gender-neutral expressions in tests designed to elicit gender stereotypes, we will randomize two versions of the tests within each classroom. The model for estimating the impact of wording will resemble the one in expression (1) but will include a test fixed effect and an interaction between the treatment and test indicators, allowing for variation in wording effectiveness between treated and control students.

4. Conclusion

Social policies have recently prioritized gender equality in STEM fields. To achieve this goal, eliciting early interest in STEM among girls is seen as a crucial measure. Our contribution to the literature will be the evaluation of the "STEM in Genere" project, an intervention aiming to boost girls' interest in STEM and combat gender stereotypes in education. The project, funded by the University of Brescia, targets primary and lower secondary school students.

The contribution of our work will be manifold. First, thanks to the evaluation design, we will be able to provide empirical evidence about the extent of gender stereotypes in Italy in a large sample of students and about the effectiveness of the intervention. This contribution of the literature is relevant since robust available evidence so far is mostly confined to anglo-saxon countries.

Second, we will embed a survey experiment within our design, to test for the impact of neutral wording on the estimation of stereotypes endorsed by the students. The literature on survey instruments on this topic is still taking its first steps especially in languages other than English, and our contribution could help future work on Romance languages to tailor survey instruments in this research field.

Third, the intervention of this case study is interesting because it could be easily scalable in case of proven effectiveness. The originality of the treatment lies in the interactive and engaging approach taken to challenge and reshape students' perceptions of science and scientific figures. The flexible duration of sessions, from a minimum of 2 to a maximum of 4 hours, adapts to the needs of classrooms, ensuring in-depth engagement easy to manage within the school context. Additionally the teacher training component, based mostly on videos, is potentially scalable with low cost.

This said, it is important to remember some shortcomings stemming from our design, above all the limited external validity and the focus on short-term effects retrieved via questionnaires. In case of detection of positive effects, it would be advisable to expand both the period of our evaluation, including medium term effects (such as track of enrolment in 9th grade) and the test area to include a larger and diverse number of schools and territories.

References

- Barone, C., 2011. Some Things Never Change: Gender Segregation in Higher Education across Eight Nations and Three Decades. *Sociol. Educ.* 84, 157–176. <https://doi.org/10.1177/0038040711402099>
- Barone, C., Assirelli, G., 2020. Gender segregation in higher education: an empirical test of seven explanations. *High. Educ.* 79, 55–78. <https://doi.org/10.1007/s10734-019-00396-2>
- Blau, F.D., Kahn, L.M., 2017. The Gender Wage Gap: Extent, Trends, and Explanations. *J. Econ. Lit.* 55, 789–865. <https://doi.org/10.1257/jel.20160995>
- Carta, F., De Philippis, M., Rizzica, L., Viviano, E., 2023. Women, labour markets and economic growth. Banca d'Italia.
- Cheryan, S., 2012. Understanding the Paradox in Math-Related Fields: Why Do Some Gender Gaps Remain While Others Do Not? *Sex Roles* 66, 184–190. <https://doi.org/10.1007/s11199-011-0060-z>
- De Gioannis, E., Pasin, G.L., Squazzoni, F., 2023. Empowering women in STEM: a scoping review of interventions with role models. *Int. J. Sci. Educ. Part B* 13, 261–275. <https://doi.org/10.1080/21548455.2022.2162832>

- Diekman, A.B., Brown, E.R., Johnston, A.M., Clark, E.K., 2010. Seeking Congruity Between Goals and Roles: A New Look at Why Women Opt Out of Science, Technology, Engineering, and Mathematics Careers. *Psychol. Sci.* 21, 1051–1057. <https://doi.org/10.1177/0956797610377342>
- Eagly, A.H., Wood, W., 2012. Social Role Theory, in: *Handbook of Theories of Social Psychology*. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom, pp. 458–476. <https://doi.org/10.4135/9781446249222.n49>
- Eccles, J.S., 2009. Expectancies, values, and academic behaviors, in: *Achievement and Achievement Motives*.
- Finson, K.D., 2002. Drawing a Scientist: What We Do and Do Not Know After Fifty Years of Drawings. *Sch. Sci. Math.* 102, 335–345. <https://doi.org/10.1111/j.1949-8594.2002.tb18217.x>
- Guetto, R., Vergolini, L., 2017. Educational expansion without equalization: a reappraisal of the 'Effectively Maintained Inequality' hypothesis in children's choice of the upper secondary track. *Eur. Soc.* 19, 1–27. <https://doi.org/10.1080/14616696.2016.1236283>
- Guiso, L., Monte, F., Sapienza, P., Zingales, L., 2008. Culture, Gender, and Math. *Science* 320, 1164–1165. <https://doi.org/10.1126/science.1154094>
- Hooven, C., 2021. T: the story of testosterone, the hormone that dominates and divides us, First US edition. ed. Henry Holt and Company, New York, New York.
- ISTAT, 2023. *Rapporto sulla competitività dei settori produttivi*.
- Kahlenberg, S.M., Wrangham, R.W., 2010. Sex differences in chimpanzees' use of sticks as play objects resemble those of children. *Curr. Biol.* 20, R1067–R1068. <https://doi.org/10.1016/j.cub.2010.11.024>
- Kang, J., Hense, J., Scheersoi, A., Keinonen, T., 2019. Gender study on the relationships between science interest and future career perspectives. *Int. J. Sci. Educ.* 41, 80–101. <https://doi.org/10.1080/09500693.2018.1534021>
- Lent, R.W., Brown, S.D., Hackett, G., 2000. Contextual supports and barriers to career choice: A social cognitive analysis. *J. Couns. Psychol.* 47, 36–49. <https://doi.org/10.1037/0022-0167.47.1.36>
- Levanon, A., Grusky, D.B., 2016. The Persistence of Extreme Gender Segregation in the Twenty-first Century. *Am. J. Sociol.* 122, 573–619. <https://doi.org/10.1086/688628>
- Lindberg, S.M., Hyde, J.S., Petersen, J.L., Linn, M.C., 2010. New trends in gender and mathematics performance: A meta-analysis. *Psychol. Bull.* 136, 1123–1135. <https://doi.org/10.1037/a0021276>
- Lomazzi, V., 2017. Gender role attitudes in Italy: 1988–2008. A path-dependency story of traditionalism. *Eur. Soc.* 19, 370–395. <https://doi.org/10.1080/14616696.2017.1318330>
- Lueptow, L.B., Garovich, L., Lueptow, M.B., 1995. The persistence of gender stereotypes in the face of changing sex roles: Evidence contrary to the sociocultural model. *Ethol. Sociobiol.* 16, 509–530. [https://doi.org/10.1016/0162-3095\(95\)00072-0](https://doi.org/10.1016/0162-3095(95)00072-0)
- Makarova, E., Aeschlimann, B., Herzog, W., 2019. The Gender Gap in STEM Fields: The Impact of the Gender Stereotype of Math and Science on Secondary Students' Career Aspirations. *Front. Educ.* 4, 60. <https://doi.org/10.3389/feduc.2019.00060>
- Manzella, E., Barbetta, G.P., Argentin, G., n.d. *Orientare alla Scelta*. <https://doi.org/10.1257/rct.7696>
- OECD, 2021. *Education at a Glance 2021: OECD Indicators, Education at a Glance*. OECD. <https://doi.org/10.1787/b35a14e5-en>
- Schaeffer, N.C., Dykema, J., 2020. Advances in the Science of Asking Questions. *Annu. Rev. Sociol.* 46, 37–60. <https://doi.org/10.1146/annurev-soc-121919-054544>