THE ECONOMIC JOURNAL

The Economic Journal, 1–37 https://doi.org/10.1093/ej/uead019
Advance Access Publication Date: 23 February 2023

HOW EFFECTIVE ARE FEMALE ROLE MODELS IN STEERING GIRLS TOWARDS STEM? EVIDENCE FROM FRENCH HIGH SCHOOLS*

Thomas Breda, Julien Grenet, Marion Monnet and Clémentine Van Effenterre

We show in a large-scale field experiment that a brief exposure to female role models working in scientific fields affects high school students' perceptions and choices of undergraduate major. The classroom interventions reduced the prevalence of stereotypical views on jobs in science and gender differences in abilities. They also made high-achieving girls in grade 12 more likely to enrol in selective and male-dominated science, technology, engineering and mathematics programs in college. Comparing treatment effects across the 56 role model participants, we find that the most effective interventions are those that improved students' perceptions of science, technology, engineering and mathematics careers without overemphasising women's under-representation in science.

The increase in women's participation in science and engineering in the United States has levelled off in the past decade (National Science Foundation, 2017). This trend, which is common to almost all OECD countries, is a source of concern for two main reasons. First, it exacerbates gender inequality in the labour market, as science, technology, engineering and mathematics (STEM) occupations offer higher average salaries (Brown and Corcoran, 1997; Black *et al.*, 2008; Blau and Kahn, 2017) and are characterised by a smaller gender wage gap (Beede *et al.*, 2011). Second, in a context of heightened concern over a shortage of STEM workers in the advanced economies,

* Corresponding author: Julien Grenet, CNRS, Paris School of Economics, 48 boulevard Jourdan, 75014, Paris, France. Email: julien.grenet@ens.fr

This paper was received on 19 October 2021 and accepted on 21 February 2023. The Editor was Barbara Petrongolo.

The data and codes for this paper are available on the Journal repository. They were checked for their ability to reproduce the results presented in the paper. The authors were granted an exemption to publish parts of their data because access to these data is restricted. However, the authors provided the Journal with temporary access to the data, which enabled the Journal to run their codes. The codes for the parts subject to exemption are also available on the Journal repository. The restricted access data and these codes were also checked for their ability to reproduce the results presented in the paper. The replication package for this paper is available at the following address: https://doi.org/10.5281/zenodo.7588802.

We are grateful to the staff at the L'Oréal Foundation, especially Diane Baras, Aude Desanges, Margaret Johnston-Clarke, Salima Maloufi-Talhi, David McDonald and Elisa Simonpietri for their continued support to this project. We also thank the staff at the French Ministry of Education (Ministère de l'Éducation Nationale, Direction de l'Évaluation, de la Prospective et de la Performance) and at the Rectorats of Créteil, Paris and Versailles for their invaluable assistance in collecting the data. This paper greatly benefited from discussions and helpful comments from Marcella Alsan, Iris Bohnet, Scott Carrell, Clément de Chaisemartin, Bruno Crépon, Esther Duflo, Lena Edlund, Ruth Fortmann, Pauline Givord, Laurent Gobillon, Marc Gurgand, Élise Huillery, Philip Ketz, Sandra McNally, Amanda Pallais and Liam Wren-Lewis. We thank participants at the ASSA/AEA Annual Meeting in Atlanta, CEPR/IZA Annual Symposium in Labor Economics 2018 in Paris, EALE 2018 in Lyon, EEA-ESEM 2018 in Cologne, Gender Economics Workshop 2018 in Berlin, Gender and Tech Conference at Harvard, IWAEE Conference in Catanzaro and Journées LAGV 2018 in Aix-Marseille. We also thank seminar participants at Bristol, DEPP, LSE, Harvard Kennedy School, HEC Lausanne, Maastrict SBE, OECD EDU Forum, PSE, Stockholm University SOFI and Université Paris 8. We are grateful to the Institut des politiques publiques (IPP) for continuous support and to Sophie Cottet for her assistance in contacting schools. Financial support for this study was received from the Fondation L'Oréal, from the Institut des Politiques Publiques and from the Agence Nationale de la Recherche through EUR grant ANR-17-EURE-0001. The project received IRB approval at J-PAL Europe and was registered in the AEA RCT Registry with ID AEARCTR-0000903.

A previous version of this paper was circulated under the title 'Do Female Role Models Reduce the Gender Gap in Science? Evidence from French High Schools'.

this trend is likely to represent a worsening loss of talent that could reduce aggregate productivity (Weinberger, 1999; Hoogendoorn *et al.*, 2013).

The under-representation of women in these traditionally male-dominated fields can also constitute a self-fulfilling prophecy for subsequent generations, as girls have little opportunity to interact with women who work in these fields and who could inspire them. A large literature has established that exposing female students to successful or admirable women can help break this vicious circle. Most of the work to date focuses on potential role models who interact on a regular basis with the individuals they may influence, such as teachers or instructors (Bettinger and Long, 2005; Carrell *et al.*, 2010; Lim and Meer, 2017), university advisors (Canaan and Mouganie, 2021) or doctors (Riise *et al.*, 2022). Recently, however, two studies have shown that a one-off exposure to external female role models can also have significant effects on female representation in male-dominated fields of study. Porter and Serra (2020) documented a positive impact of two female role models who were carefully selected among the economics alumnae of Southern Methodist University in the United States, on the likelihood of female students majoring in economics. Del Carpio and Guadalupe (2021) demonstrated the effectiveness, relative to other types of intervention, of a virtual role model in reducing identity costs related to female participation in STEM and fostering female applications to a software-coding program.[1] An attractive feature of these light-touch interventions for identifying role model effects is that they remove the influences of potential confounding factors such as gender differences in teaching practices (Lavy and Sand, 2018; Carlana, 2019; Terrier, 2020).

While the studies by Porter and Serra (2020) and Del Carpio and Guadalupe (2021) furnish compelling evidence that external role models can affect female students' educational choices, little is known about what drives their success, and it is unclear whether different role models are equally able to influence students' decisions. This paper addresses these questions by evaluating the impact of one-hour in-class interventions by women scientists. Our key contribution is to characterise what makes for an effective role model intervention. We investigate the attributes of role models and the messages they convey that are more likely to appeal to young women's perceptions, and to trigger their interest in traditionally male-dominated fields. Two distinctive aspects of our setting make it particularly well suited to address this question. First, we use a large-scale randomised experiment on a diverse student population, exploiting both a rich post-intervention survey and comprehensive administrative data to measure directly how role models affect students' perceptions, beliefs and enrolment outcomes. Second, unlike previous studies, our research design involves a large number of role model participants—56 in all. We leverage the diversity of these women's profiles to better understand what makes an effective role model.

The program we evaluate is called 'For Girls in Science', launched in 2014 by the L'Oréal Foundation—the corporate foundation of the world's leading cosmetics manufacturer—to encourage girls to explore STEM career paths. It consists of one-hour in-class interventions by women with two quite distinct profiles: half are young scientists (either PhD candidates or postdoctoral researchers) who were awarded the L'Oréal-UNESCO Fellowship 'For Women in Science'; the others are young professionals employed as scientists in the Research and Innovation division of the L'Oréal group. In the main part of the intervention, the role models share their experience and career path with the students. They also provide information on science-related careers in general and on gender stereotypes, using two short videos.

[1] Related studies outside the context of STEM education include field experiments on exposure to women in leadership positions in India (Beaman *et al.*, 2012), and the provision of information on the returns to education by role models of poor or rich background in Madagascar (Nguyen, 2008).

The evaluation was conducted during the 2015/16 academic year in 98 high schools in the Paris region. It involved 19,451 students from grade 10 and grade 12 (science track), two grade levels at the end of which students make irreversible educational choices. Half of the classes were randomly selected to be visited by one of the 56 role model participants, who were assigned to those classes through a registration process on a first-come, first-served basis.

The role models' interventions led to a significant increase in the share of girls enrolling in STEM fields, but only in the educational tracks where they are severely under-represented. In grade 10, the classroom visits had no detectable impact on boys' and girls' probability of enrolling in the science track in grade 11, where girls are only slightly under-represented (47% of students). In grade 12, by contrast, the intervention induced a significant increase in the share of female students enrolling in selective STEM undergraduate programs, which lead to the most prestigious graduate schools, and in male-dominated STEM programs (maths, physics, computer science and engineering).[2] The visits respectively increased enrolment by 3.1 and 3.4 percentage points (pps) in selective and male-dominated programs among girls in grade 12, or increases of 28% and 20% over the baseline rates of 11% and 17%. These effects are concentrated among high-achieving girls in maths. Although we cannot formally reject the equality of gender coefficients, the effects for boys are small in magnitude and not statistically significant. These results constitute the first field evidence that in-person exposure to external female role models directly influences STEM enrolment decisions at college entry.

To explore the channels through which role models affect students' enrolment outcomes, we conducted a post-treatment student survey consisting of an eight-page questionnaire administered in class between one and six months after the classroom interventions. We also collected administrative data on high school graduation exams (*baccalauréat*) at the end of grade 12. Our results show that the role model interventions significantly improved students' perceptions of science-related jobs at both grade levels, with no indication of declining effects over a period of up to six months. They also helped mitigate some of the stereotypes typically associated with STEM occupations (such as the difficulty of reconciling them with family life) and heightened the perception that these jobs pay better. By contrast, the interventions had no significant effect on students' self-reported taste for science subjects or their academic performance, and bolstered the girls' self-concept in maths only slightly, at either grade level.

One of the most interesting—and unexpected—findings concerns the effects on students' perceptions of gender roles in science. Not only were the classroom interventions effective in debiasing students' beliefs about gender differences in maths aptitude, they also raised awareness of the under-representation of women in science. The combination of these two effects triggered an unintended ex post rationalisation by students of the gender imbalance in scientific fields and occupations, making them more likely to agree with the statements that women dislike science and that they face discrimination in science-related jobs. Explicitly correcting self-stereotyping beliefs (Coffman, 2014) and misperceptions about women's representation in science (Bursztyn and Yang, 2022) would thus appear to have generated more ambiguous perceptions among students than the intervention's gender-neutral messages about jobs and careers.

[2] We classify STEM programs as being male dominated if the share of female students is less than 50% (see Online Appendix A for details). Note that male-dominated and selective STEM programs are partly overlapping: in 2016/17, 49% of undergraduate students in male-dominated STEM fields were enrolled in selective programs, while 95% of students in selective STEM programs were in male-dominated fields.

Finally, we highlight the importance of the role model's profile for the success of the intervention. We document a high degree of heterogeneity in treatment effects according to the role model's professional background. Those employed by the sponsoring firm had a significantly greater effect on girls' probability of enrolling in selective STEM programs than the young researchers, even though the two sets of students they visited had similar observable characteristics. While the two groups of role models were equally effective in debunking the stereotype on gender differences in maths aptitude, we find clear evidence that those working at L'Oréal were more effective in improving girls' perceptions of science-related jobs and elevating their aspirations for such careers. Conversely, they were less likely to reinforce students' beliefs that women are under-represented in science. Using machine learning methods, we provide further evidence that the most effective role models are those who managed to convey a positive image of science careers, and to stimulate students' aspirations without overemphasising the relative scarcity of women and its possible causes. Together, these results show that role model interventions are not reducible to the provision of standardised information and that female role models are not interchangeable. They also highlight the mechanisms that likely explain the substantial effects that have been documented in other settings where career women serve as external role models (Porter and Serra, 2020).

The remainder of the paper is organised as follows. Section 1 provides institutional background on the French educational system and the gender gap in STEM fields. Section 2 describes the intervention and the experimental design. Section 3 presents the data and the empirical strategy. Section 4 analyses the effects of role model interventions on student perceptions, self-concept, aspirations and educational outcomes. Section 5 extends the analysis to the role of information, the persistence of effects and potential spillovers. Section 6 discusses what makes an effective role model intervention and Section 7 concludes.

1. Institutional Background

1.1. *Structure of the French Education System*

In France, education is compulsory from 6 to 16. The school year runs from September to June. The school system consists of five years of elementary education (grades 1 to 5) and seven years of secondary education, divided into four years of middle school (*collège*, grades 6 to 9) and three of high school (*lycée*, grades 10 to 12). Students complete high school with the national *baccalauréat* exam, which they must pass for admission to higher education.

1.1.1. *High school tracks*

The tracking of students occurs at two critical stages (see Figure 1). At the end of middle school, about two-thirds of students are admitted to general and technical upper secondary education (*seconde générale et technologique*) and the remaining third are tracked into vocational schools (*seconde professionnelle*). After the first year of high school (grade 10), the general and technical tracks are further split: approximately 80% of the students are directed to the general *baccalauréat* program for the last two years of high school (grades 11 and 12), and the other 20%, mostly low-achieving students, are directed towards a technical *baccalauréat*, which is more geared towards the needs of business and industry and leads to shorter studies.

In the spring term of grade 10, the students who have been allowed to pursue the general track are required to choose among three sub-tracks in grade 11: science (*Première S*), humanities

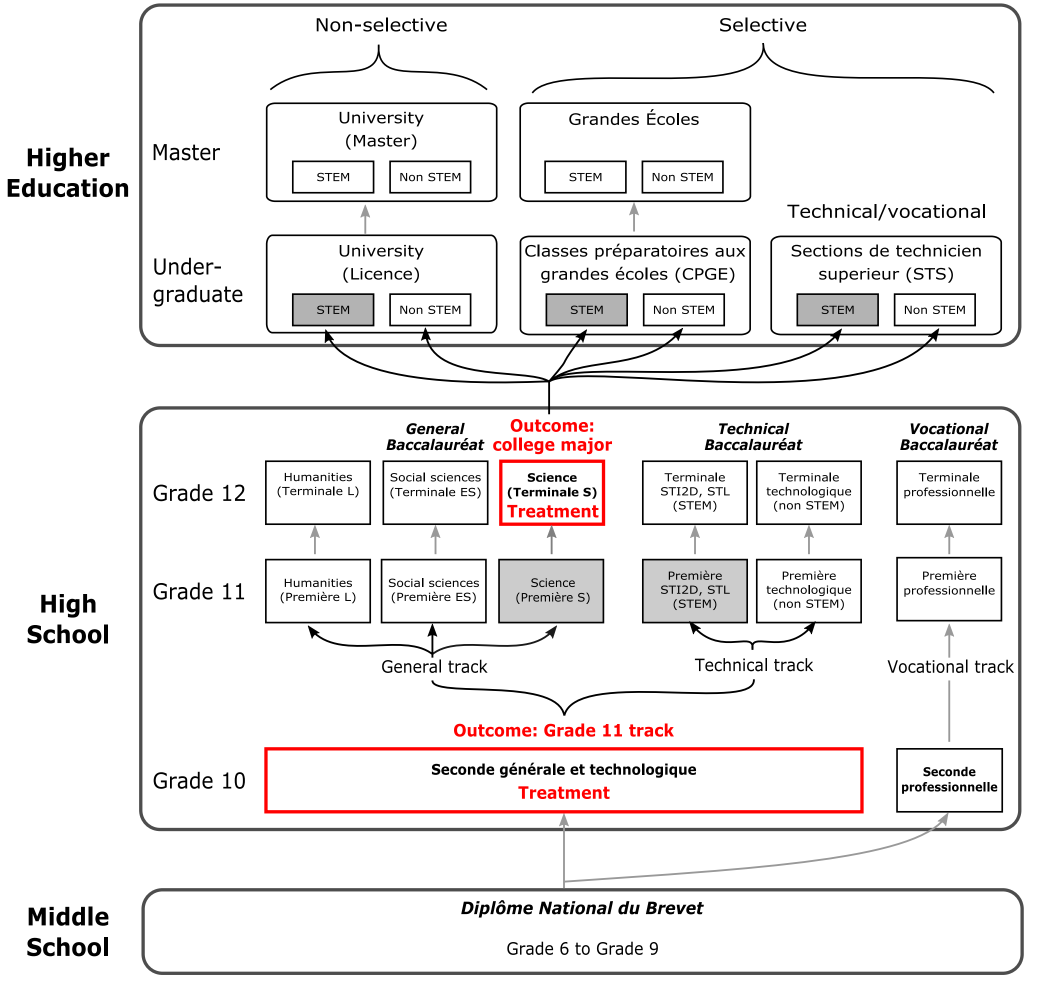

Fig. 1. *Tracks in Secondary and Post-Secondary Education in France.*

(*Première L*) and social sciences (*Première ES*). This is an important choice, given that the curriculum and high school examinations are specific to each *baccalauréat* track, and thus have a direct impact on students' educational opportunities and career prospects. It is almost impossible, for instance, for a student to be admitted to engineering or medical undergraduate programs without a *baccalauréat* in science. Students directed to the technical track after grade 10 are also required to choose among eight possible STEM and non-STEM sub-tracks, which will affect their field of study in higher education.

1.1.2. *College entry*

In the spring term of grade 12, students in their final year of high school apply for admission to higher education programs through a centralised online admission platform. The programs to which students can apply fall into two broad categories, each accounting for about half of first-year undergraduate enrolment: (*i*) non-selective undergraduate university programs (*licence*), which are open to all students who hold the *baccalauréat*; and (*ii*) selective programs, which can admit or reject students based on their academic achievement. Both types of program offer specialisations in STEM and non-STEM fields. The most prestigious selective programs are the two-year *classes préparatoires aux grandes écoles* (CPGE), which prepare students to take the national entry exams to elite graduate schools (*grandes écoles*). These programs are specialised either in science, in economics and business or in humanities. Within the science CPGE programs, the main fields of specialisation are mathematics and physics (MPSI), physics and chemistry (PCSI) and biology/geoscience (BCPST). The other selective undergraduate programs (*section de technicien supérieur* or STS) are mostly targeted to students holding a vocational or technical *baccalauréat* and prepare for technical/vocational bachelor's degrees.

1.2. *Female Under-Representation in STEM*

In France, the share of female students in STEM-oriented studies starts to decline after grade 10 and drops sharply at entry into higher education. While 54% of the students in the general and technical track in grade 10 are girls, the share falls to 47% in the general science track (grades 11 and 12) and then plummets to 30% in the first year of higher education.[3] Female under-representation in STEM fields of study is more pronounced in the selective undergraduate programs (shares of 18% in STS and 30% in CPGE) than in the non-selective programs (35%). These proportions, which are derived from administrative data for 2016/17, are almost identical to those of a decade earlier. Within STEM fields, female students tend to specialise in earth and life sciences (female share: 62%) rather than mathematics, physics or computer science (female share: 26%).

The under-representation of women in STEM fields accounts for a good part of the gender pay gap among university graduates in France. Using a variety of administrative and survey data sources (MESRI-DGESIP/DGRI-SIES, 2017; CGE, 2018; MESRI, 2018), we show that across all majors, male graduates who obtained a master's degree in 2015 or 2016 earn a median gross annual starting salary of €32,122, compared to €28,411 for female graduates (Online Appendix A). This gap of €3,711 per year is equal to 11.6% of men's pay (see Online Appendix Table A1). Using standard decomposition methods, we find that the under-representation of female students in STEM accounts for approximately 28% of this gap (see Online Appendix Table A2). Additionally, almost half of the 9.1% gender pay gap within STEM can be ascribed to the fact that female graduates are less likely than males to be enrolled in the selective and male-dominated fields, which lead to the best-paying degrees. These figures strongly suggest that, in the French context, increasing the share of female students in STEM—especially in selective and male-dominated programs—would narrow the gender pay gap substantially.

2. Program and Experimental Design

2.1. *'For Girls in Science'*

The program 'For Girls in Science' (FGiS) is an awareness campaign launched in 2014 by the L'Oréal Foundation to encourage girls to explore STEM career paths. It consists of one-

[3] At the high school level, the gender imbalance in STEM is much more severe in the technical track (female share: 17%) than in the general science track (female share: 47%).

hour one-off classroom interventions by female role models with a background in science. The interventions, which take place in the presence of all students in the class, including boys, are made by female role models of two distinct types: (*i*) PhD candidates or postdoctoral researchers who have been awarded a fellowship by the Foundation (the L'Oréal-UNESCO 'For Women in Science' Fellowship) and who participate in the program as part of their contract; and (*ii*) young scientists employed in the research and innovation division of the L'Oréal group who volunteer for the program.

2.1.1. *Structure and content of the interventions*

The classroom interventions last one hour and are divided into four main sequences. Each role model was given a set of slides as a support for the in-class conversation. During the first sequence, a few slides highlight two facts: (1) the labour market is marked by high demand for STEM skills and there is a shortage of graduates in the relevant fields of study; and (2) women are under-represented in STEM careers. To investigate the role of information provision, we gave 36 of the 56 role models additional slides that they were free to use during this sequence. They supplied supplementary information about average earnings and employment conditions in STEM jobs, and were illustrated with examples of career prospects in humanities versus science. In Section 5, we discuss the sensitivity of our results to this more intensive provision of standardised information.

The second sequence kicks off with two three-minute videos designed to set forth and deconstruct stereotypes about science-related careers and gender roles in science.[4] The first video, entitled 'Science, Beliefs or Reality?', uses interviews with high school students to debunk myths about careers in science (e.g., jobs in science are more challenging, they necessarily require more years of schooling), stereotypes about scientists (e.g., they are introverted, lonely) and gender differences in science aptitude (e.g., women are naturally less talented in maths). The second video, entitled 'Are we all Equal in Science?', sets out the common gender stereotypes about science aptitude while providing information on brain plasticity, and on how interactions and the social environment shape men's and women's abilities and tastes. This sequence seeks to stimulate class discussion based on students' reactions.

The third sequence centres on the role model's own experience as a woman with a background in science and consists of a question-and-answer session with the students.[5] Topics addressed during this discussion include the role model's typical day at work, what she enjoys about her job, the biggest challenge she had to overcome, how she views her professional future, her everyday interactions with co-workers, how much she earns and her work-family balance. Consistent with the program's emphasis on the role model dimension, this sequence was intended as the longest and most important part of the intervention. To convey this objective to the role models, a day-long training session was organised to help them share their experience with the students. The training also included a workshop on the under-representation of women in science and a practice session aimed at enhancing oral communication skills.

The intervention concludes with an overview of the diversity of STEM studies and careers, illustrated by concrete examples such as jobs in graphic design, environmental engineering and computer science.

[4] Screenshots of the two videos shown during the classroom interventions are displayed in Online Appendix Figure B1.
[5] Screenshots of the slides used during the discussion are shown in Online Appendix Figure B2.

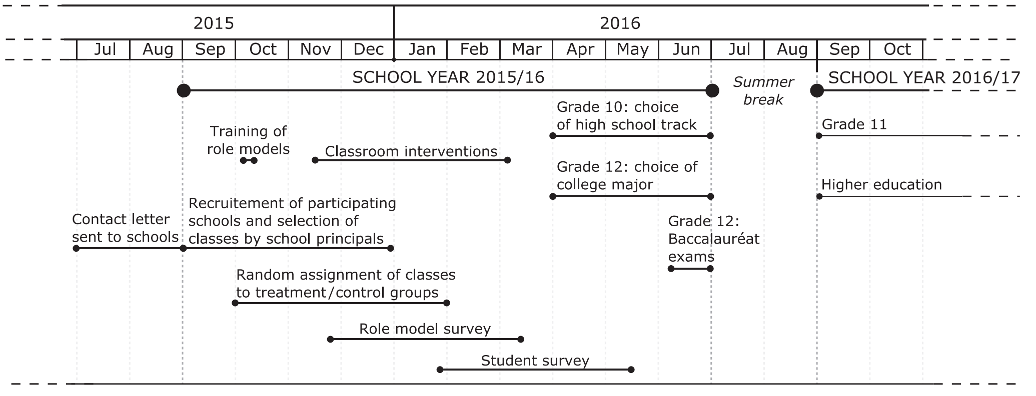

Fig. 2. *Program Evaluation Timeline.*

2.2. *Experimental Design*

2.2.1. *Participating schools*

The evaluation was conducted in the three school districts (*académies*) of the Paris region (Paris, Créteil and Versailles) during the 2015/16 academic year. Créteil and Versailles are the two largest districts in France, and the three combined include 318,000 high school students in the general and technical track, or 20% of all French high school enrolment.

Figure 2 shows the detailed timeline of the evaluation. In the spring of 2015, the French Ministry for Education agreed to support a randomised evaluation of the program, and designated one representative for each district as intermediary between the schools and the evaluation team. In June, official letters informed the high school principals, who are in charge of extracurricular activities, that they were likely to be contacted to take part in the evaluation. All public and private high schools with at least four classes in grade 10 and two in the grade 12 science track, were contacted by our team between September and December 2015, accounting for 349 of the 489 high schools in the three districts. Of these schools, 98 agreed to take part in the experiment, representing 28% of grade 10 enrolment and 29% of grade 12 science track enrolment in the three districts.[6] The participating schools tend to be larger and are less likely to be private or to be in the Paris education district than the non-participants (see Online Appendix Table E1).

2.2.2. *Selection of classes and randomisation*

In the fall of 2015, the principals were invited to select at least six classes—four or more in grade 10 and two or more in grade 12 science track—and to indicate a preferred time slot and day for the visits.[7] While it is possible that principals selected classes where they expected the interventions to be more effective, Online Appendix Table E2 shows that selected and non-selected classes are broadly comparable, lending support to the external validity of our experimental design. The gender composition of grade 10 classes is similar between the two groups, while in grade 12, the share of female students is slightly higher in the classes selected. Despite these differences, the experimental sample, which consists of 19,451 students (13,700 in grade 10 and 5,751 in

[6] The location of the participating schools is shown in Online Appendix Figure B3. The high schools that declined to take part did so mainly because they feared the organisational burden of the classroom interventions.

[7] In a large majority of schools, principals selected exactly four grade 10 and two grade 12 classes.

grade 12), resembles the relevant student population quite closely, both in social composition and in average academic performance.

In each school, half of the classes selected (up to the nearest integer) were assigned randomly to the treatment group (302 classes in total) and the other half to the control group (299 classes). Table 1 indicates that, while the random assignment successfully balanced the characteristics of students in the treatment and control groups in grade 10, it did not achieve perfect covariate balance in grade 12. In our empirical analyses, we account for these residual imbalances by controlling for students' baseline characteristics in our main specification.

2.2.3. *Role models*

The experiment involved 56 female role models: 35 L'Oréal employees and 21 PhD candidates or postdoctoral researchers. Table 2 provides summary statistics of their characteristics. The researchers are younger (30 versus 36 years of age on average) and less likely to be foreign nationals (10% versus 17%). Although both types have very high educational attainment, 39% having graduated from a *grande école*, the researchers, by definition, are more likely than the professionals to hold (or be preparing for) a PhD (100 versus 38%) and to hold a degree in maths, physics and engineering (38% versus 14%). They are also less likely to have children (19% versus 58%) and to have been involved in the program in the previous year (19% versus 29%). The professionals working at L'Oréal are employed in various activities: chemistry (development of new technologies for skin products), logistics and supply chain management, statistics (consumer evaluation), immunology and toxicology. Although we could not collect direct information on earnings for reasons of confidentiality, based on aggregate information provided by the L'Oréal Group, we estimate that the annual gross salary of these young professionals is between €45,000 and €65,000, compared with €22,000–€50,000 for the researchers. On average, the role models carried out five classroom interventions in two different high schools.

2.2.4. *Classroom interventions*

The classroom visits took place between 17 November 2015 and 3 March 2016.[8] The role models were asked to choose two or three schools in which to make an average of three classroom visits per school—in most cases, two in grade 10 and one in grade 12. They were not assigned to the schools randomly, but registered for the visits and time slots on a first-come, first-served basis during four registration sessions using an online system. Randomly assigning the role models to the schools was not feasible, as most were participating on a voluntary basis and during regular working hours. We therefore gauge the causal impact of role models in a setting where they have some freedom to choose the schools in which they intervene. The assignment process, however, did not involve any coordination between the participants, and was designed to limit their ability to select the schools they would visit, as each registration session only concerned a subset of the participating schools.[9]

[8] Of the visits, 17% took place in November, 26% in December, 40% in January, 17% in February and 1% in March (see panel A of Online Appendix Table E9).

[9] The role models were contacted four times to complete the schedule, on 21 October, 24 November, 7 December 2015 and 3 February 2016.

Table 1. *Treatment-Control Balance.*

	Control group (1)	Treatment group (2)	Within school: Difference T − C (3)	Within school: *p*-value of diff. (4)
Panel A. Grade 10				
Student characteristics				
Female	0.535	0.522	−0.010	0.309
Age (years)	15.13	15.12	−0.01	0.180
Non-French	0.059	0.061	0.002	0.652
High SES	0.377	0.386	0.008	0.321
Medium-high SES	0.131	0.125	−0.007	0.168
Medium-low SES	0.248	0.235	−0.012	0.064
Low SES	0.244	0.254	0.012	0.085
Number of siblings	1.485	1.486	0.003	0.904
Class size	33.22	33.27	0.07	0.476
At least one science elective course	0.389	0.398	0.005	0.820
At least one standard elective course	0.770	0.737	−0.031	0.138
DNB percentile rank in maths	58.61	58.35	−0.35	0.533
DNB percentile rank in French	57.79	57.91	0.12	0.829
Test of joint significance	*F*-statistic: 0.798 (*p*-value: 0.653)			
Predicted track in grade 11				
Grade 11: science track	0.449	0.452	0.001	0.922
Grade 11: science—general track	0.373	0.375	0.001	0.920
Grade 11: science—technical track	0.077	0.077	−0.000	0.989
N	6,801	6,899	13,700	
Panel B. Grade 12 (science track)				
Student characteristics				
Female	0.499	0.484	−0.014	0.292
Age (years)	17.14	17.11	−0.04	0.000
Non-French	0.053	0.048	−0.006	0.275
High SES	0.453	0.474	0.029	0.009
Medium-high SES	0.136	0.135	−0.001	0.829
Medium-low SES	0.216	0.201	−0.015	0.023
Low SES	0.195	0.190	−0.012	0.140
Number of siblings	1.510	1.487	−0.032	0.127
Class size	31.75	32.19	0.39	0.196
DNB percentile rank in maths	74.17	73.95	0.20	0.699
DNB percentile rank in French	69.31	69.90	0.89	0.122
Test of joint significance	*F*-statistic: 0.983 (*p*-value: 0.459)			
Predicted undergraduate major				
Major: STEM	0.382	0.384	0.003	0.352
Major: selective STEM	0.175	0.178	0.006	0.081
Major: male-dominated STEM	0.273	0.276	0.004	0.279
N	2,853	2,898	5,751	

Notes: Each row corresponds to a different linear regression with the dependent variable listed on the left, separately for students in grade 10 (panel A) and in grade 12 (panel B). Columns (1) and (2) show the average value for students in the control and treatment groups, respectively. Column (3) reports the coefficient from the regression of each variable on the treatment group indicator, with the *p*-value reported in column (4). The regression controls for school fixed effects to account for the fact that randomisation was stratified by school, and standard errors are adjusted for clustering at the unit of randomisation (class). The *F*-statistic is from a test of the joint significance of the coefficients in a regression of the treatment group indicator on all student characteristics. High school tracks (panel A) and undergraduate majors (panel B) are predicted for each student using the coefficients from a linear regression of the corresponding binary variable (e.g., enrolment in a STEM major) on all student characteristics listed in the table. This model is fitted separately by grade level on the sample of students in the control group.

Table 2. *Female Role Models: Summary Statistics.*

	All role models (1)	Researchers (PhD/postdoc) (2)	Professionals (employed by sponsoring firm) (3)
Age ($N = 51$)	33.3	30.0	35.6
	(5.7)	(3.1)	(6.0)
Non-French	0.14	0.10	0.17
Holds/prepares for a PhD ($N = 55$)	0.62	1.00	0.38
Graduated from a grande école	0.39	0.33	0.43
Field: maths, physics, engineering	0.23	0.38	0.14
Field: earth and life sciences	0.64	0.62	0.66
Field: other	0.13	0.00	0.20
Has children ($N = 52$)	0.42	0.19	0.58
Participated in the program the year before	0.25	0.19	0.29
Number of high schools visited	1.8	2.1	1.6
	(0.8)	(0.9)	(0.7)
Number of classroom interventions	5.2	5.9	4.7
	(2.3)	(2.3)	(2.1)
Visited at least one high school in Paris	0.27	0.29	0.26
N	56	21	35

Notes: The summary statistics are computed based on information obtained from the L'Oréal Foundation and from the post-intervention survey administered online to collect feedback about the classroom visits. SDs are shown in parentheses below the mean values. Where data are missing for some role models, the number of non-missing values N is indicated in parentheses.

3. Data and Empirical Strategy

3.1. *Data*

To evaluate the program's effects on student perceptions and educational outcomes, we combine three main data sources: (*i*) a post-intervention survey of the role models; (*ii*) a post-intervention survey of the students; and (*iii*) student-level administrative data.[10]

3.1.1. *Role model survey*

After each visit to a school, the role models were invited to complete an online survey (Breda *et al.*, 2016a,b). Besides collecting general feedback, this survey served to monitor compliance with random assignment, asking them to indicate each of the classes they visited. Summary statistics are reported in Online Appendix Table E3. The interventions almost always (89%) took place in the presence of the teacher and sometimes (35%) of another adult. The role models reported organisational problems (e.g., the intervention started late, the slides could not be shown) for only 16% of the visits. According to the survey, researchers and professionals were equally likely to cover the intended topics, such as 'jobs in science are fulfilling', 'they are for girls too' and 'they pay well'. Finally, when asked about their overall perception of their interventions, 93% gave positive assessments, saying they went 'well' (37%) or 'very well' (56%). Students were generally seen to have been responsive to the key messages.

3.1.2. *Student survey*

We conducted a paper-and-pencil student survey in the classes assigned to the treatment and control groups between one and six months after the classroom visits, i.e., between January

[10] The original and translated versions of the two surveys are provided in the replication package to this paper.

and May 2016 (Breda *et al.*, 2016c). Each questionnaire had a unique identifier so that it could be linked with student-level administrative data. The survey was designed to collect a rich set of information on students' preferences, beliefs and perceptions regarding science, self-concept and aspirations. The questionnaire was anonymous and, to maximise the response rate and the quality of the responses, was administered in exam conditions under the supervision of a teacher (not necessarily the one present during the classroom visit). It was presented as a general survey on attitudes about science and science-related careers, so as to minimise the risk that students would associate it with the FGiS program and, therefore, reduce the scope for social desirability bias.[11] It was eight pages long and took about half an hour to complete.

The survey items are designed to measure the effects of the interventions on students' perceptions along five dimensions: (*i*) general perceptions of science-related careers; (*ii*) perceptions of gender roles in science; (*iii*) taste for science subjects; (*iv*) self-concept in maths; and (*v*) science-related career aspirations. When conceptually related, we combine the survey items to construct a composite index for each dimension using standardised *z*-score scales. Section 4 below describes the specific items that are used for each dimension.[12]

As shown in Online Appendix Table E5, the survey response rates are high both in grade 10 (88%) and in grade 12 (91%). They are slightly higher among grade 10 students in the treatment than in the control group (by 2.6 pps). Despite this small difference in response rates, the characteristics of survey respondents are generally balanced (see Online Appendix Table E6). In analysing survey-based outcomes, we control for students' characteristics at baseline to account for any residual imbalance between the treatment and control groups.

3.1.3. *Administrative data*

We linked the student survey data to a rich set of individual-level administrative data covering the universe of high school students in the Paris region from 2012/13 to 2016/17 (DAPEP, 2017; PAPP, 2017; SSA, 2017). These data provide detailed information on students' socio-demographic characteristics and enrolment status every year, allowing us to identify the track taken by grade 10 students entering grade 11.

The college enrolment outcomes of students in grade 12 were obtained by matching the survey and administrative data for high school students with administrative microdata, covering almost all the students enrolled in selective and non-selective higher education programs in 2016/17 (MESRI-DGESIP/DGRI-SIES, 2017).[13] These data are supplemented by comprehensive individual examination results from the *diplôme national du brevet* (DNB), which is taken at the end of middle school, and from the national *baccalauréat* exam for grade 12 students (MENJ-DEPP, 2017). Specifically, we use students' grades on the final exams in French and maths (converted into national percentile ranks), as these tests are graded externally and anonymously. Further details on the data sources and the classification of higher education programs can be found in Online Appendix C.

[11] In Section 5, we provide suggestive evidence against experimenter demand effects driving our findings.

[12] To attenuate potential order bias, the order of several of the response items (e.g., maths/French, man/woman) was set randomly.

[13] The programs not covered by these administrative data are those leading to paramedical and social care qualifications. Available estimates suggest that among grade 12 students who obtained a *baccalauréat* in science in 2008, under 6% were enrolled in those programs the following year (Lemaire, 2012).

3.2. *Empirical Strategy*

Compliance with random assignment was not perfect: about 5% of the classes assigned to the treatment group were not visited by a role model, and 1% of the classes in the control group, instead, were mistakenly visited (see Online Appendix Table E4).[14] To deal with this marginal two-way non-compliance, we follow the standard practice of using treatment assignment as an instrument for treatment receipt, which allows us to estimate the program's local average treatment effect (LATE) instead of the average treatment effect. Specifically, we estimate the following model using two-stage least squares:[15]

$$Y_{ics} = \alpha + \beta D_{cs} + \boldsymbol{X}_{ics}\pi + \theta_s + \epsilon_{ics}, \quad (1)$$

$$D_{ics} = \gamma + \delta T_{cs} + \boldsymbol{X}_{ics}\tau + \lambda_s + \eta_{ics}. \quad (2)$$

Here Y_{ics} denotes the outcome of student i in class c and high school s, D_{cs} is a dummy variable indicating whether the student's class received a visit and T_{cs} is a dummy for assignment to the treatment group. The regression further includes the student characteristics $\boldsymbol{X}_{ics}$ listed in Table 1 to control for residual imbalances between the treatment and control groups. Finally, school fixed effects, θ_s and λ_s, are included to account for the fact that the randomisation was stratified by school and grade level.

The model specified by (1) and (2) is estimated separately by grade level and gender, with standard errors clustered at the unit of randomisation (class). To account for multiple hypothesis testing across the outcomes of interest, the treatment effect estimates are accompanied by adjusted p-values (q-values) in addition to the standard p-values.[16]

4. Effects of Classroom Interventions

We analyse the impact of the classroom interventions on three sets of student outcomes: (*i*) general perceptions of science-related careers and of gender roles in science; (*ii*) preferences, self-concept and aspirations; and (*iii*) enrolment outcomes and academic performance.

4.1. *Perceptions of STEM Careers and Gender Roles in Science*

Students' post-intervention survey responses show that the classroom interventions were effective in challenging stereotyped views of science-related careers and gender roles. The results are reported in Table 3 for students in grade 10 and in Table 4 for students in grade 12.

[14] We are confident that non-compliance was mostly due to organisational and logistical issues and was not an endogenous response to randomisation. The few role models who carried out interventions in classes assigned to the control group, or in classes not selected to participate in the evaluation generally reported that their interventions, had been poorly organised, the person in charge often not being aware of the purpose of the visit. In some cases, classroom interventions were scheduled during another speciality course involving multiple classes, meaning that only some of the students in the treatment group were effectively treated.

[15] Because non-compliance concerned only a small fraction of classes, the LATE and intention-to-treat (ITT) estimates are very close in magnitude. The ITT estimates can be found in Online Appendix Table H1 (columns (1) and (4)).

[16] We use the false discovery rate (FDR) control, which designates the expected proportion of all rejections that are type-I errors. Specifically, we use the sharpened two-stage q-values introduced in Benjamini *et al.* (2006) and described in Anderson (2008). See Online Appendix D for details.

Table 3. *Impact of Role Model Interventions on Student Perceptions in Grade 10.*

	Girls			Boys			
	Control group mean (1)	Treatment effect (LATE) (2)	*p*-value [*q*-value] (3)	Control group mean (4)	Treatment effect (LATE) (5)	*p*-value [*q*-value] (6)	*p*-value of diff. (5) − (2) (7)
Panel A. Perceptions of science-related careers							
Positive perceptions of science-related careers (index)	−0.020	0.245*** (0.027)	0.000 [0.001]	0.023	0.162*** (0.027)	0.000 [0.001]	0.013
Panel B. Perceptions of gender roles in science							
More men in science-related jobs	0.628	0.154*** (0.013)	0.000 [0.001]	0.629	0.170*** (0.014)	0.000 [0.001]	0.345
Equal gender aptitude for maths (index)	0.115	0.111*** (0.024)	0.000 [0.001]	−0.134	0.142*** (0.030)	0.000 [0.001]	0.383
Women do not really like science	0.157	0.056*** (0.011)	0.000 [0.001]	0.198	0.101*** (0.013)	0.000 [0.001]	0.002
Women face discrimination in science-related jobs	0.603	0.126*** (0.013)	0.000 [0.001]	0.527	0.154*** (0.014)	0.000 [0.001]	0.102
Panel C. Stated preferences, self-concept and aspirations							
Taste for science subjects (index)	−0.169	−0.033 (0.031)	0.275 [0.414]	0.197	−0.021 (0.026)	0.431 [0.555]	0.704
Self-concept in maths (index)	−0.198	−0.001 (0.028)	0.981 [0.982]	0.231	0.033 (0.029)	0.250 [0.375]	0.324
Science-related career aspirations (index)	−0.103	0.005 (0.029)	0.851 [0.970]	0.120	0.004 (0.027)	0.871 [0.872]	0.977
N		6,475			5,751		

Notes: This table reports estimates of the treatment effects of the role model interventions on the perceptions of students in grade 10. The sample is restricted to students who completed the post-intervention questionnaire. Each row corresponds to a different linear regression performed separately by gender, with the dependent variable listed on the left. Columns (1) and (4) report the average value for students in the control group. Columns (2) and (5) report the LATE estimates. They are obtained from a regression of the outcome of interest on a classroom visit indicator, using treatment assignment as an instrument for treatment receipt. The regression includes school fixed effects (to account for the fact that randomisation was stratified by school) and the student characteristics listed in Table 1. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomisation (class). Columns (3) and (6) report the cluster-robust *p*-value of the estimated treatment effect and, in square brackets, the *p*-value (*q*-value) adjusted for multiple hypothesis testing, using the false discovery rate (FDR) control method. Specifically, we use the sharpened two-stage *q*-values introduced in Benjamini *et al.* (2006) and described in Anderson (2008). The *q*-values are adjusted for multiple testing across the study's nine main outcomes of interest, separately by grade level and gender (see Online Appendix D for details). The *p*-value of the difference between the treatment effects by gender is reported in column (7). $^{***}p < 0.01$.

Table 4. *Impact of Role Model Interventions on Student Perceptions in Grade 12 (Science Track).*

	Girls			Boys			
	Control group mean (1)	Treatment effect (LATE) (2)	p-value [q-value] (3)	Control group mean (4)	Treatment effect (LATE) (5)	p-value [q-value] (6)	p-value of diff. (5) − (2) (7)
Panel A. Perceptions of science-related careers							
Positive perceptions of science-related careers (index)	−0.003	0.296*** (0.032)	0.000 [0.001]	0.003	0.171*** (0.033)	0.000 [0.001]	0.002
Panel B. Perceptions of gender roles in science							
More men in science-related jobs	0.712	0.122*** (0.016)	0.000 [0.001]	0.717	0.149*** (0.015)	0.000 [0.001]	0.166
Equal gender aptitude for maths (index)	0.158	0.078*** (0.028)	0.004 [0.007]	−0.161	0.124*** (0.042)	0.003 [0.006]	0.348
Women do not really like science	0.074	0.042*** (0.009)	0.000 [0.001]	0.146	0.073*** (0.015)	0.000 [0.001]	0.079
Women face discrimination in science-related jobs	0.624	0.085*** (0.020)	0.000 [0.001]	0.600	0.074*** (0.018)	0.000 [0.001]	0.651
Panel C. Stated preferences, self-concept and aspirations							
Taste for science subjects (index)	−0.002	0.018 (0.033)	0.583 [0.583]	0.002	0.014 (0.040)	0.733 [0.825]	0.924
Self-concept in maths (index)	−0.184	0.051 (0.035)	0.139 [0.157]	0.187	0.068** (0.033)	0.038 [0.057]	0.695
Science-related career aspirations (index)	−0.045	0.106*** (0.037)	0.004 [0.007]	0.046	0.068* (0.035)	0.055 [0.071]	0.410
N		2,600			2,636		

Notes: This table reports estimates of the treatment effects of the role model interventions on the perceptions of students in grade 12 (science track). The sample is restricted to students who completed the post-intervention questionnaire. Each row corresponds to a different linear regression performed separately by gender, with the dependent variable listed on the left. Columns (1) and (4) report the average value for students in the control group. Columns (2) and (5) report the LATE estimates. They are obtained from a regression of the outcome of interest on a classroom visit indicator, using treatment assignment as an instrument for treatment receipt. The regression includes school fixed effects (to account for the fact that randomisation was stratified by school) and the student characteristics listed in Table 1. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomisation (class). Columns (3) and (6) report the cluster-robust p-value of the estimated treatment effect and, in square brackets, the p-value (q-value) adjusted for multiple hypothesis testing, using the FDR control method. Specifically, we use the sharpened two-stage q-values introduced in Benjamini *et al.* (2006) and described in Anderson (2008). The q-values are adjusted for multiple testing across the study's nine main outcomes of interest, separately by grade level and gender (see Online Appendix D for details). The p-value of the difference between the treatment effects by gender is reported in column (7). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

4.1.1. *Perceptions of science-related careers*

Students were asked to agree or disagree with five statements on science-related careers relating to pay, the length of studies leading to these careers, work-life balance and the two prevalent stereotypes that science-related jobs are monotonous and solitary. We build a composite index of 'positive perceptions of science-related careers' by re-coding the Likert scales, so that higher values correspond to less stereotyped or negative perceptions, before taking the average of each student's responses to the five questions. To facilitate interpretation, we normalise the index to have a mean of zero and an SD of 1 in the control group.[17] For closer investigation of the various aspects that might be captured by the overall index, we construct binary variables taking value 1 if the student agrees (strongly or somewhat) with each statement, and zero if he/she disagrees.[18]

One of the interventions' key objectives was to correct students' beliefs about jobs and careers in science by offering, not only standardised information, but also information specific to each role model's experience. As is shown in panel A of Tables 3 and 4, the role model interventions significantly improved girls' and boys' perceptions of such careers as measured by the composite index. The treatment effect estimates range from 16% of an SD for boys to around 30% for girls, with significantly stronger effects for female students in both grades.

The detailed results for the different components of the index are reported in Online Appendix Table F1. Students' baseline perceptions indicate relatively widespread negative stereotypes about careers in science (see columns (1) and (4)), with little difference between boys and girls or between grade levels. As an example, between 17% and 33% of students consider that science-related jobs are monotonous or solitary. A significant impact of the classroom visits is observed for almost all the components of the index. The largest effects relate to the statements 'science-related jobs require more years of schooling' and 'science-related jobs are rather solitary', two stereotypes that were explicitly debunked in the slides and videos. Although the effects are not strikingly different between genders and grade levels, they tend to be greater for girls in grade 12. In particular, the interventions significantly reinforced female students' perceptions that science-related careers are compatible with a fulfilling family life, a message specifically conveyed by the role models, and in line with the evidence showing that jobs in science and technology enable women to work more flexibly (Goldin, 2014).

4.1.2. *Perceptions of gender roles in science*

Female under-representation in STEM can be broadly attributed to three possible causes: gender differences in abilities, discrimination (on the demand side) and differences in preferences and career choices (on the supply side). The survey questions were designed to capture students' views on these three dimensions.[19]

Strikingly, the results show that more than a third of grade 10 students and a quarter of grade 12 students in the control group are not aware that women are under-represented in science-related

[17] We checked that our results are robust to converting the item responses into binary variables before computing the indices, as well as to using Bartlett factor scores instead of the procedure described in the text. See Online Appendix D for further details on the construction of the composite indices.

[18] Similar groupings are performed when using responses that are measured on a four-point Likert scale (usually concerning perceptions or self-confidence) so that the outcome variables can be directly interpreted as proportions. The results are not qualitatively affected by such grouping.

[19] Unlike the survey questions related to students' perceptions of science-related careers, those on perceptions of gender roles in science are not aggregated into a single index, because they were designed to capture different dimensions that cannot be easily combined. For the same reason, we refrained from using a single index to measure students' stated preferences, self-concept and aspirations (see Section 4.2).

careers (panel B of Tables 3 and 4). These proportions do not differ greatly either by gender or by grade. For boys and girls in both grades, the interventions increased awareness of female under-representation in STEM by 12 to 17 pps. This is one of the strongest effects of the interventions.

The classroom interventions were also effective in debiasing students' beliefs about gender differences in maths aptitude. To capture this dimension, we asked students whether they agreed with the statements that 'men are more gifted than women in mathematics' and that 'men and women are born with different brains'. We used these two questions to construct a composite index to gauge whether students believe that men and women have equal aptitude for mathematics. The results show significant rises in this index for both genders in both grades, with treatment effects ranging between 7.8% and 14.2% of an SD.[20]

Interestingly, the classroom visits had more ambiguous, partially unintended effects regarding the other two possible causes. First, when asked about gender differences in preferences, the share of students who agree that 'women do not really like science' is relatively low in the control group (16% of girls and 20% of boys in grade 10; 7% of girls and 15% of boys in grade 12), but it is substantially higher owing to the interventions for both genders, by 4 to 10 pps. Second, the baseline shares of boys and girls who say that women face discrimination in science-related jobs are much larger (between 53% and 62%); these too increase for both genders, by 7 to 15 pps. These unintended effects on students' perceptions might represent an effort to rationalise the small number of women in science-related careers, making students more likely to agree with the simplistic view that 'women do not really like science', as well as subscribing to the idea that women face discrimination.

4.2. *Stated Preferences, Self-Concept and Aspirations*

We now turn to the effects of the interventions on students' tastes for science subjects, their self-concept in maths and their science-related career aspirations. The results are reported in panel C of Tables 3 and 4.

4.2.1. *Taste for science subjects*

For both genders in grade 10 and grade 12, the classroom visits had no sizeable impact on students' taste for science subjects, which we measure using an index that combines their answers to four questions about their enjoyment of maths, physics-chemistry, and earth and life sciences (on a 0 to 10 Likert scale), and their self-reported taste for science in general (on a four-point Likert scale).[21] These findings are not particularly surprising, given that the interventions did not expose students to science-related content and were not specifically designed to promote interest in science.

4.2.2. *Self-concept in maths*

To measure the impact of the classroom visits on students' self-concept in mathematics, we use a composite index combining the responses to four questions: (*i*) students' self-assessed performance in maths; (*ii*) whether they feel lost when trying to solve a maths problem; (*iii*) whether they often worry that they will struggle in maths class; and (*iv*) whether they think they can do well in science subjects if they make enough effort.

[20] The detailed results for the two components of this index are reported in Online Appendix Table F2.
[21] The detailed results for the four components of the index are reported in Online Appendix Table F3.

Consistent with the literature, our sample exhibits large gender differences in self-concept in mathematics. In the control group, the value of the index is 43% of an SD lower for girls than for boys in grade 10, and 37% lower in grade 12. Large gender differences are found for most of the items used in the construction of this index, in particular those related to maths anxiety (see Online Appendix Table F4).

Although the interventions were light touch, they did have some positive effect on students' self-concept in maths. These effects are statistically significant only for boys in grade 12 when using the composite index. The interventions, however, consistently reduced the probability of students reporting worry that they will struggle in maths class. Point estimates tend to be higher for boys than for girls in both grades, indicating that the classroom interventions had no corrective effect on the substantial gender gap in this area.

4.2.3. *Science-related career aspirations*

The choice of a science-related career path does not depend solely on students' tastes for the science taught at school. It also depends on their perceptions of the relevant jobs and their amenities, such as earnings, work/life balance and the work environment, all of which were embodied by the role models.

To measure the effects on students' aspirations for science-related careers, we use a composite index combining the responses to four questions: (*i*) whether the students find that some jobs in science are interesting; (*ii*) whether they could see themselves working in a science-related job; (*iii*) whether they are interested in at least one of six STEM jobs out of a list of 10 STEM and non-STEM occupations[22]; and (*iv*) whether they consider career and earnings prospects as important factors in their choice of study.

Although the interventions had no discernable impact on grade 10 students' science-related aspirations, in grade 12 the effects are positive and statistically significant for both genders (11% of an SD for girls, significant at the 1% level, and 7% for boys, significant at the 10% level). The more detailed results reported in Online Appendix Table F5 show that the interventions had significant positive effects on three of the relevant survey items for grade 12 students. In particular, those in the treatment group are more likely to report that career and earnings prospects are important factors in their choice of study, which is consistent with the thesis that the interventions raised their awareness of the wage premium for STEM jobs.

4.3. *Educational Choices and Academic Performance*

4.3.1. *High school track after grade 10*

Table 5 shows that the classroom visits had no significant impact on grade 10 students' choices of track for the academic year following the intervention, i.e., 2016/17. For both genders, the treatment effect estimates are close to zero, whether we consider enrolment in any STEM track or in the general and technical STEM tracks separately. That is, the interventions had no effect on the 20 pp gender gap in the probability of pursuing STEM studies after grade 10.

Several mechanisms can be posited for this lack of effect on the enrolment status of grade 10 girls the next year. First, the interventions did not appear to be especially well suited to increase the share of girls enrolling in the STEM technical tracks in grade 11, where their share is particularly low (17%). As is noted below, the positive effects observed on the STEM enrolment

[22] The STEM occupations in the list were chemist, computer scientist, engineer, industrial designer, renewable energy technician and researcher in biology. The non-STEM occupations were lawyer, pharmacist, physician and psychologist.

Table 5. *Impact of Role Model Interventions on Grade 10 Students' Enrolment Status the Following Year.*

	Girls			Boys			
	Control group mean (1)	Treatment effect (LATE) (2)	p-value [q-value] (3)	Control group mean (4)	Treatment effect (LATE) (5)	p-value [q-value] (6)	p-value of diff. (5) − (2) (7)
All STEM tracks							
Grade 11: science track	0.355	−0.002	0.862	0.551	−0.006	0.640	0.800
		(0.011)	[0.970]		(0.012)	[0.720]	
General versus technical STEM track							
Grade 11: science—general track	0.328	0.003	0.794	0.416	0.004	0.710	0.925
		(0.010)	[0.794]		(0.011)	[0.710]	
Grade 11: science—technical track	0.026	−0.005	0.188	0.135	−0.010	0.234	0.562
		(0.004)	[0.377]		(0.008)	[0.468]	
Other tracks or repeater							
Grade 11: other tracks	0.545	0.006	0.614	0.324	0.020	0.110	0.391
		(0.012)			(0.012)		
Repeater or dropout	0.101	−0.004	0.636	0.126	−0.014*	0.090	0.327
		(0.009)			(0.008)		
N		7,241			6,459		

Notes: This table reports estimates of the treatment effects of classroom interventions on grade 10 students' enrolment outcomes in the academic year following the classroom interventions, i.e., 2016/17. The enrolment outcomes are measured using student-level administrative data. Each row corresponds to a different linear regression performed separately by gender, with the dependent variable listed on the left. Columns (1) and (4) report the average value for students in the control group. Columns (2) and (5) report the LATE estimates. They are obtained from a regression of the outcome of interest on a classroom visit indicator, using treatment assignment as an instrument for treatment receipt. The regression includes school fixed effects (to account for the fact that randomisation was stratified by school) and the student characteristics listed in Table 1. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomisation (class). Columns (3) and (6) report the cluster-robust p-value of the estimated treatment effect and, in square brackets, the p-value (q-value) adjusted for multiple hypothesis testing, using the FDR control method. Specifically, we use the sharpened two-stage q-values introduced in Benjamini *et al.* (2006) and described in Anderson (2008). The q-values associated with the treatment effect estimates on 'Grade 11: science track' are adjusted for multiple testing across the study's nine main outcomes of interest, separately by gender (see Online Appendix D for details). The q-values associated with the treatment effect estimates on enrolment in the general and technical STEM tracks are adjusted for multiple testing across these two tracks, separately by gender. The p-value of the difference between the treatment effects by gender is reported in column (7). $^{*}p < 0.1$.

decisions of girls in grade 12 are concentrated among the high achievers in maths. In grade 10, high-achieving students are traditionally not directed to the technical tracks. If low-achieving students are less likely to be affected, then it is not surprising to find limited effects on enrolment in STEM technical tracks. Turning to the general science track, female under-representation is only moderate in grade 11 (in 2016/17, the female share was 47%), and this track is the most common choice (usually the default) for high-performing students, including girls. Unlike the other high school tracks, it gives access to almost all fields of study in higher education, and thus does not signal any strong commitment to a STEM education or career in the future; that is, the potential of STEM role models to influence enrolment in this track is limited. Female students who turn away from the science track in high school are unlikely to even consider a STEM career as a viable option, making their choices less easily reversible.[23]

4.3.2. *Field of study after grade 12*

A central finding of the study is that the role model interventions had significant effects on the educational choices of girls in grade 12, by raising their probability of enrolling in selective and in male-dominated STEM programs in higher education.[24]

Table 6 indicates a positive but statistically insignificant ($p = 0.14$) effect of the interventions on the probability of female students enrolling in undergraduate STEM programs, of 2.0 pps from a baseline of 28.9%, i.e., a 7% increase. Importantly, however, we find that the classroom visits had larger and statistically significant effects on female students' enrolment in the STEM programs in which they are most severely under-represented. Our estimates show that their probability of enrolling in selective STEM programs increased by 3.1 pps (a 28% increase from the baseline of 11.0%, significant at the 1% level). In male-dominated STEM programs (mathematics, physics, computer science and engineering), their enrolment probability increased by 3.4 pps from a baseline of 16.6% (i.e., a 20% increase, significant at the 1% level).

These results are particularly striking, given that selective and male-dominated STEM programs are, not only the most prestigious tracks, but also those where the gender gap is greatest, explaining approximately half of the STEM-related gender pay gap in France (see the discussion in Section 1.2). Our estimates indicate that, on average, the role model interventions induced one girl in every two grade 12 science-track classes to switch to a selective or a male-dominated STEM program at entry into higher education.[25]

The more detailed results presented in Online Appendix Table F6 suggest that these effects are driven by girls switching from non-STEM programs and from STEM programs that are neither selective nor male dominated. A significant decline in female enrolment is indeed found for non-selective undergraduate programs in earth and life sciences (−2.2 pps), while small reductions of 0.2 to 0.9 pp are found for selective programs in humanities and non-STEM vocational programs, as well as for non-selective programs in medicine, law and economics, humanities and psychology, and sports studies. Role models thus appear to have affected the enrolment outcomes of grade 12 girls who would have otherwise chosen a curriculum in a female-dominated environment, be it in STEM or outside STEM.

[23] Consistent with this interpretation, the survey data indicate that among grade 10 students in the control group, only 24% of the girls who did not enrol in the science track the following year said they could see themselves working in a science-related job, compared to 87% of those who did.

[24] STEM programs are classified as being either male dominated or female dominated, depending on whether the share of female students in the corresponding field is below or above 50% (see Online Appendix C.2 for details).

[25] This calculation is based on an average of 15 girls per class.

Table 6. *Impact of Role Model Interventions on Grade 12 Students' Enrolment Status the Following Year.*

	Girls			Boys			
	Control group mean (1)	Treatment effect (LATE) (2)	p-value [q-value] (3)	Control group mean (4)	Treatment effect (LATE) (5)	p-value [q-value] (6)	p-value of diff. (5) − (2) (7)
All undergraduate STEM majors							
Major: STEM	0.289	0.020	0.139	0.470	−0.002	0.925	0.310
		(0.014)	[0.157]		(0.019)	[0.926]	
Selective versus non-selective STEM							
Major: selective STEM	0.110	0.031***	0.006	0.232	0.008	0.575	0.200
		(0.011)	[0.012]		(0.015)	[0.575]	
Major: non-selective STEM	0.178	−0.011	0.333	0.239	−0.010	0.445	0.959
		(0.012)	[0.333]		(0.013)	[0.575]	
Male- versus female-dominated STEM							
Major: male-dominated STEM	0.166	0.034***	0.004	0.379	0.013	0.485	0.289
(maths, physics, computer science)		(0.012)	[0.012]		(0.019)	[0.575]	
Major: female-dominated STEM	0.123	−0.015	0.169	0.091	−0.015	0.119	0.983
(earth and life sciences)		(0.011)	[0.226]		(0.009)	[0.477]	
Other tracks or dropout							
Other non-STEM programs	0.507	−0.031**	0.049	0.293	−0.008	0.571	0.286
		(0.016)			(0.015)		
Not enrolled in higher education	0.206	0.011	0.430	0.237	0.012	0.425	0.957
		(0.013)			(0.015)		
N		2,827			2,924		

Notes: This table reports estimates of the treatment effects of classroom interventions on grade 12 (science track) students' enrolment outcomes in the academic year following the classroom interventions, i.e., 2016/17. The enrolment outcomes are measured using student-level administrative data. Each row corresponds to a different linear regression performed separately by gender, with the dependent variable listed on the left. Columns (1) and (4) report the average value for students in the control group. Columns (2) and (5) report the LATE estimates. They are obtained from a regression of the outcome of interest on a classroom visit indicator, using treatment assignment as an instrument for treatment receipt. The regression includes school fixed effects (to account for the fact that randomisation was stratified by school) and the student characteristics listed in Table 1. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomisation (class). Columns (3) and (6) report the cluster-robust p-value of the estimated treatment effect and, in square brackets, the p-value (q-value) adjusted for multiple hypothesis testing, using the FDR control method. Specifically, we use the sharpened two-stage q-values introduced in Benjamini *et al.* (2006) and described in Anderson (2008). The q-values associated with the treatment effect estimates on 'Major: STEM' are adjusted for multiple testing across the study's nine main outcomes of interest, separately by gender (see Online Appendix D for details). The q-values associated with the treatment effect estimates on enrolment in the different STEM majors are adjusted for multiple testing across these different majors, separately by gender. The p-value of the difference between the treatment effects by gender is reported in column (7). ***$p < 0.01$, **$p < 0.05$.

By contrast, we find no evidence of statistically significant effects of the interventions on the college major decisions of boys in grade 12. The estimated effect on their probability of enrolling in a STEM undergraduate program is close to zero (−0.2 pp from a baseline of 47.0%), while the effects on enrolment in selective STEM (0.8 pp from a baseline of 23.2%) and male-dominated STEM programs (1.3 pp from a baseline of 37.9%) are small and insignificant at conventional levels. It should be noted, however, that these estimates do not allow drawing firm conclusions on the impact of the program on the gender gap in STEM enrolment, as we lack the statistical power to reject the null hypothesis of equal effects on male and female students.

Taken together, the results for grade 12 students indicate that the interventions were effective in steering girls towards the STEM tracks in which they are heavily under-represented, even though two-thirds of the role models came from female-dominated STEM fields (earth and life sciences), and that the interventions were designed to promote all types of STEM careers, including those where women now outnumber men. These findings suggest that in the current setting, the role models affected only the most strongly stereotyped choices.

4.3.3. *Academic performance*

The effects of the classroom visits on academic performance can be documented for students in grade 12 based on the *baccalauréat* exams, taken a few months after the classroom interventions. The estimates of the effect of the treatment on students' performance on the maths test and on the probability of obtaining the *baccalauréat* are close to zero and statistically insignificant for both genders (see Online Appendix Table F7).[26] Although hypothetically the role models could have strengthened students' motivation to be admitted to the most selective STEM programs, and so increased the time devoted to studying maths and science, we find no evidence of any such effect. We can therefore rule out that the effects on the enrolment outcomes of girls in grade 12 were driven by increased effort and accordingly better academic performance.

4.4. *Robustness Checks*

We conducted a series of robustness checks for our main findings (see Online Appendices G and H).

First, we checked that our results are robust to using a specification that does not control for students' characteristics at baseline. The resulting estimates for the survey-based outcomes (Online Appendix Table G1) are quite similar to those presented in Tables 3 and 4. The estimates without controls are also qualitatively similar to those reported in Tables 5 and 6 for enrolment outcomes (Online Appendix Table G2). They tend to be slightly larger (but not statistically significant) for boys in grade 12, which we interpret as a consequence of the small residual imbalances in the male sample.[27]

Second, we assessed the sensitivity of our results to non-parametric randomisation inference tests rather than model-based cluster-robust inference. The tests are performed by comparing our ITT estimates with the distribution of 'placebo' ITT estimates obtained by randomly re-assigning treatment 2,000 times among participating classes within each school and grade level.

[26] The small negative effect on overall *baccalauréat* performance for female students is only marginally significant when we control for student characteristics and is not robust to omitting these controls.

[27] Balancing tests performed separately by grade level and gender do not point to unusually large covariate imbalance between the treatment and control groups in any of the subsamples (results available upon request). However, the predicted probability of being enrolled in a selective STEM program is marginally higher in the treatment than in the control group for boys in grade 12 (by 0.8 pp from a baseline of 23.8%, significant at the 5% level).

The results yield empirical p-values that are generally close to the model-based p-values (see Online Appendix Table H1). Although they tend to be slightly more conservative, they confirm the interventions' statistically significant effects on female enrolment in selective and male-dominated undergraduate STEM programs.

5. Information, Persistence and Spillovers

In this section, we test the sensitivity of students' attitudes and choices to the informational component of the intervention. We then extend the analysis to the persistence of effects on student perceptions, the timing of the interventions and the potential spillover effects on enrolment outcomes.

5.1. *The Role of Information Provision*

Role model interventions, not only foster self-identification, but intrinsically contain an informational component. While our design does not allow fully disentangling these two mechanisms, there is suggestive evidence that the purely informational component of the classroom visits does not in itself explain the changes in female students' college major decisions after grade 12.

As described in Section 2, we initially sent a set of slides to the role models to assist them during the intervention. The first six slides highlighted some stylised facts about jobs in science and female under-representation in STEM careers, but gave only limited information on employment conditions in such careers, and no information on salaries. Starting on 20 November 2015, we sent six additional slides to 36 of the 56 role models, with more detailed information regarding wage and employment gaps between STEM and non-STEM jobs, as well as differences between male and female students' choices of study. The role models were free to integrate these slides into their final presentation or just use them as a support.[28]

The results reported in Online Appendix I show that students' characteristics are balanced between the role models who received the standard or the 'augmented' set of slides (see Online Appendix Table I1).[29] Consistent with the thesis that the effects on college major decisions were not driven primarily by the standardised information contained in the slides, we find that the role models who had just the standard slideshow also had positive and significant effects on the probability of female students enrolling in selective STEM and male-dominated STEM programs after grade 12 (see Online Appendix Table I2). And we find no evidence that those who had the additional slides had significantly larger effects on girls' STEM enrolment outcomes, although the students with whom they interacted were more likely to agree that science-related jobs pay higher salaries.[30]

[28] Screenshots of the two sets of slides are shown in Online Appendix Figures I1 and I2.

[29] At first, we planned to allocate the two sets of slides randomly to the role models and were able to do so for a subset of 14 participants. However, the L'Oréal Foundation requested that, going forward, all remaining role models be provided with the 'augmented' version. Those who had already started the visits kept the standard version. To ensure sufficient statistical power, we present results for the entire sample of role models, controlling for month-of-visit fixed effects. The results are qualitatively similar if we restrict the sample to the subset to whom the slides were randomly assigned.

[30] Note that, since the use of the supplementary slides was at the discretion of the role models, the coefficient on the interaction term $T \times$ additional slides in Online Appendix Table I2 should be interpreted as a lower bound for the effect of the more information-intensive treatment.

5.2. *Persistence*

The effects on students' perceptions that we observe could be short lived. We explore this issue by comparing the treatment effects, depending on the time elapsed between the classroom visit and the date when the student completed the survey. Splitting the sample at the median of this time interval (63 days), we find that, on average, students below this threshold completed the survey 46 days after the intervention, those above it in 93 days, i.e., an extra 47 days. Note that students whose class was visited early are more likely to have waited longer before completing the survey. The comparison of treatment effects between the two subsamples should therefore be interpreted with some caution when assessing persistence, since these effects may also capture heterogeneity related to the timing of the visits (see the next paragraph). Moreover, the interval between the intervention and survey completion never exceeds six months. With these caveats in mind, the results in Online Appendix Table I3 suggest that the treatment effects did not vanish quickly, insofar as they are statistically significant and of comparable magnitude in both subsamples and, in most cases, are not significantly different. These results should also attenuate concerns about social desirability bias, since experimenter demand effects would be expected to be greater for students who took the survey shortly after the intervention.

5.3. *Timing of Visits*

We find suggestive evidence that earlier interventions had greater effects on the college choices of grade 12 students, which could be made through May. For girls, the positive effects on enrolment in STEM, selective STEM and male-dominated STEM, are all statistically significant for the classroom visits that took place in November or December 2015, whereas the effects of visits in January or February 2016 are smaller and not significant (see Online Appendix Table I4).[31] With the caveat that we cannot reject the null hypothesis of equal treatment effects across the two subperiods, these findings suggest that interventions made when many students are still undecided about their field of study and career plans, may be more effective than those on the eve of the deadline when irreversible choices may already have been made.

5.4. *Spillovers*

An important issue is whether the interventions could have influenced the educational choices of students in the control group. These students may have heard about the visits directly, through their schoolmates in treatment group classes, or indirectly, through regular social interaction. If the direction of such effects is the same for students in the treatment and control groups, ignoring spillovers would cause us to underestimate the treatment effects.

On the last page of the post-intervention survey questionnaire, the students in the treatment group were asked whether they had discussed the classroom intervention with their classmates, with schoolmates from other classes or with friends outside of school, as a way of assessing possible spillover effects. Students in the control group received a slightly different version of this final section, asking whether they had heard of classroom visits by male or female scientists in other classes, with no explicit mention of the FGiS program.

[31] We are confident that these differences are not driven primarily by confounding factors. Although the subsamples defined on the basis of whether the visit took place in 2015, or in 2016, exhibit significant imbalances with respect to education districts and the share of private schools, they are reasonably balanced with respect to students' characteristics and predicted STEM enrolment (results available upon request).

The survey evidence suggests that the scope for spillover effects was limited, which is consistent with the idea that in French schools most peer interactions take place within the class (Avvisati *et al.*, 2014). In the treatment group, 58% of grade 10 students and 63% of grade 12 students report having talked about the classroom intervention with their classmates, but only 24% and 27% report having talked with schoolmates from other classes (see Online Appendix Table J1). In the control group, only 14% of students in grade 10 report having heard of the classroom visits, almost all of them (12%) only vaguely. In grade 12, students in the control group are more likely (34%) to report being at least vaguely aware of the visits, but fewer than 5% of boys and girls have a precise recollection. Overall, these summary statistics suggest that spillover effects were quite limited indeed.

We complement this survey evidence with a more formal investigation of whether the interventions affected the higher education choices of grade 12 students whose classes were not assigned to the treatment group—either classes not selected by principals for the interventions or participating classes randomly assigned to the control group. Our empirical strategy, described in detail in Online Appendix J, builds on the following intuition: for schools that participated in the evaluation, the random assignment of treatment to participating classes makes it possible to estimate the average outcome that would have resulted if *all* students had only been exposed to the spillover effects of classroom interventions without being *directly* exposed to a role model. This unobserved 'spillover-only' counterfactual can be estimated at the school level by computing an appropriately weighted average of the outcomes of students in the non-participating classes and in the participating classes that were assigned to the control group. Students in the control group classes are given a greater weight, as they are used to account for both their own outcome and for the hypothetical outcome in the treatment classes, if they had been exposed to a role model only indirectly.[32] The spillover effects of the interventions are then estimated by comparing the 'spillover-only' counterfactual and a 'no-treatment' counterfactual. This second counterfactual is constructed using non-participating schools, which we observe in the administrative data, whose observable characteristics are similar to those of the participating schools over the period 2012–5. Having verified that trends in student enrolment outcomes were parallel between the two groups of schools in the pre-treatment period, we implement a difference-in-differences estimator to identify the interventions' spillover effects on students' STEM enrolment outcomes at college entry.

This difference-in-differences approach produces no evidence of significant spillover effects on non-treated grade 12 students (see Online Appendix Table J2). Together with the survey evidence, the results based on this approach suggest that spillovers between treatment and control classes were at most limited.

6. What Makes the Role Model Intervention Effective?

To understand what drives the success of the interventions, we investigate the characteristics of the message, the messenger and the students who were the most responsive. One advantage of

[32] For instance, in a school with two participating classes, one treated and one control, and one non-participating class, the 'spillover-only' counterfactual is computed by assigning a weight of 1 to the non-participating class and a weight of 2 to the control group class (if all classes have the same number of students). By virtue of randomisation, mean outcomes in the control classes provide unbiased estimates of the counterfactual 'spillover-only' outcomes in the treatment classes.

our setting is that we can compare treatment effects for groups of students who were exposed to different role models or who responded differently to the same one.

We proceed in three steps. First, we show that the treatment effects on STEM enrolment outcomes vary substantially along the two most salient dimensions of heterogeneity, namely the role models' background (L'Oréal professionals versus researchers) and the students' academic performance. Second, we determine which of the student perceptions were most strongly affected by the role models who had the greatest impact on enrolment outcomes. Third, we build on the machine learning approach of Chernozhukov *et al.* (2018), to analyse whether the students who were particularly receptive or unreceptive to some of the messages conveyed are the same ones whose choice of study was most or least affected by the interventions. We use this approach to determine which messages were most effective.

6.1. *Heterogeneous Treatment Effects on STEM Enrolment*

We start by investigating how the treatment effects on STEM enrolment vary with the role models' background and the students' performance in maths. Our analysis focuses on grade 12 students, as we find no evidence of significant effects on enrolment outcomes for grade 10 students.[33]

6.1.1. *Role model background: researchers versus professionals*

We find clear evidence that the two types of role model had different effects on the STEM enrolment outcomes of girls in grade 12 (see panel A of Table 7 and Online Appendix Figure K1). The professionals increased the probability of female students enrolling in a selective STEM program by a significant 5.4 pps, whereas the researchers had no discernable effect.[34] The contrast is qualitatively similar whether male-dominated STEM programs or all STEM programs are considered. While the estimates also point to larger effects for boys who were exposed to role models with a professional background, they are not statistically significant at conventional levels.

Why were the two types of role model not equally able to steer female students towards STEM fields? The academic role models are, on average, younger than the professionals employed by the sponsoring firm (see Table 2), which might foster greater identification on the part of the students. But because they work in highly specialised fields and in very competitive environments, it is not clear how attainable students might think their achievements are. On the other hand, the professionals tend to have higher pay and more experience, and they come less often from a purely academic background. Also, unlike PhD candidates and postdocs, they hold permanent jobs and their work environment could be perceived as more attractive. Finally, the types of role model might differ in their communication skills and charisma.[35] While it is hard to pinpoint the

[33] The results of the heterogeneity analysis by role model background and maths performance for grade 10 students are reported in Online Appendix Tables K1 and K2.

[34] The difference between the treatment effects of the two types of role model is significant at the 5% level.

[35] Although we cannot rule this explanation out, we do not think it is the most likely, both because researchers and professionals received a one-day training before visiting the high schools, and because the PhD candidates and postdocs, with their experience as teaching assistants, are probably more used to speaking to a student audience and handling classrooms. It is also possible that the professionals were more motivated than the researchers because they volunteered for the program. Our feeling is that this aspect may not have played a major role. We met the academic role models on multiple occasions, and our general impression is that they were genuinely enthusiastic about their participation.

Table 7. *Heterogeneous Treatment Effects on Grade 12 Students' Outcomes, by Role Model Background.*

	Girls			Boys		
	Role model background			Role model background		
	Researchers (1)	Professionals (2)	*p*-value of diff. [*q*-value] (3)	Researchers (4)	Professionals (5)	*p*-value of diff. [*q*-value] (6)
Panel A. Enrolment outcomes						
Undergraduate major: STEM	−0.014	0.043***	0.049	−0.017	0.010	0.480
	(0.024)	(0.016)	[0.146]	(0.030)	(0.024)	[0.729]
Undergraduate major: selective STEM	−0.002	0.054***	0.017	−0.016	0.028	0.151
	(0.019)	(0.013)	[0.034]	(0.024)	(0.019)	[0.303]
Undergraduate major: male-dominated STEM	0.015	0.046***	0.188	−0.008	0.029	0.340
	(0.019)	(0.015)	[0.189]	(0.028)	(0.026)	[0.340]
N	1,180	1,647		1,312	1,612	
Panel B. Student perceptions						
Positive perceptions of science-related careers (index)	0.166***	0.380***	0.001	0.157***	0.181***	0.711
	(0.049)	(0.039)	[0.003]	(0.046)	(0.046)	[0.800]
More men in science-related jobs	0.137***	0.113***	0.461	0.161***	0.139***	0.486
	(0.025)	(0.021)	[0.649]	(0.023)	(0.021)	[0.729]
Equal gender aptitude for maths (index)	0.095**	0.068**	0.623	0.185***	0.078	0.208
	(0.046)	(0.033)	[0.702]	(0.067)	(0.053)	[0.729]
Women do not really like science	0.039***	0.044***	0.815	0.091***	0.060***	0.284
	(0.015)	(0.012)	[0.816]	(0.025)	(0.017)	[0.729]
Women face discrimination in science-related jobs	0.124***	0.061***	0.135	0.088***	0.064***	0.509
	(0.035)	(0.024)	[0.243]	(0.028)	(0.023)	[0.729]
Taste for science subjects (index)	−0.054	0.065	0.074	−0.011	0.033	0.567
	(0.049)	(0.044)	[0.167]	(0.056)	(0.055)	[0.729]
Self-concept in maths (index)	0.081	0.032	0.504	0.126**	0.023	0.118
	(0.059)	(0.043)	[0.649]	(0.050)	(0.043)	[0.729]
Science-related career aspirations (index)	−0.089	0.231***	0.000	0.060	0.074	0.846
	(0.057)	(0.044)	[0.001]	(0.054)	(0.047)	[0.847]
N	1,067	1,533		1,174	1,462	

Notes: This table reports estimates of the treatment effects of the role model interventions on the outcomes of grade 12 students, separately by gender and by background of the female role model who visited the classroom (professional or researcher). Each row corresponds to a different linear regression performed separately by gender, with the dependent variable listed on the left. Columns (1) and (2) (for girls) and columns (4) and (5) (for boys) report the LATE estimates for students whose class was visited by a researcher or a professional, respectively. They are obtained from a regression of the outcome of interest on the interaction between a classroom visit indicator and indicators for the role model being either a researcher or a professional, using treatment assignment (interacted with the role model background indicator) as an instrument for treatment receipt. The regression includes school fixed effects (to account for the fact that randomisation was stratified by school) and the student characteristics listed in Table 1. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomisation (class). Columns (3) and (6) report both the cluster-robust model-based *p*-value for the difference between the treatment effect estimates for students visited by a professional versus a researcher and, in square brackets, the *p*-value (*q*-value) adjusted for multiple hypothesis testing, using the FDR control method. Specifically, we use the sharpened two-stage *q*-values introduced in Benjamini *et al.* (2006) and described in Anderson (2008). $^{***}p < 0.01$, $^{**}p < 0.05$.

precise attributes that could explain why the professionals had more impact than the researchers, our data allow us to investigate the messages they conveyed more effectively to the students (see Section 6.2).

6.1.2. *High versus low achievers in maths*

Academic performance in mathematics is the single most important admission criterion of selective undergraduate STEM programs. Using grade 12 students' national percentile rank on the *baccalauréat* maths test to proxy for academic performance, we find that the interventions' positive impact on selective STEM enrolment is driven by female students above the median (see panel A of Table 8).[36] For these girls, the probability of enrolling in a selective STEM program after high school increases by 6.5 pps (significant at the 1% level), which corresponds to a 34% increase from the baseline of 19%, while the effect is close to zero for girls below the median and is not statistically significant for boys. The differences in treatment effects between high- and low-achieving girls in maths are qualitatively similar for enrolment in male-dominated STEM programs and for all STEM programs.[37]

6.1.3. *Potential confounders*

Even though the role models were not randomly assigned to the participating schools, the classroom visits of the researchers and the professionals are similarly distributed over the period of intervention (see Online Appendix Table E9, panel A).[38] The characteristics of the schools and students visited by the two sets of role models also appear to be reasonably balanced (see Online Appendix Tables E7 and E8). There are, however, a few statistically significant differences. In grade 12, in particular, the professionals were more likely than the researchers to visit private high schools (24% versus 10%).

Despite these small imbalances, Table 9 shows that the significantly larger impact of professionals on selective STEM enrolment for grade 12 girls is robust to controlling for a full set of interactions between the treatment group dummy and the observable characteristics of students and schools (columns (1) and (2)), as well as for interactions between the treatment dummy and the role models' characteristics and the month of intervention (column (3)). That is, there is no indication that the heterogeneous treatment effects according to the role models' background are confounded by differential selection into schools or by other observable characteristics of the role models.[39] Table 9 further shows that the larger treatment effects for high-achieving girls in maths are robust to controlling for the same set of interactions.[40]

[36] As noted in Section 4.3, we find no significant impact of the interventions on students' performance on the maths test of the *baccalauréat* exam, which mitigates concerns about potential endogenous selection bias when conditioning on this variable.

[37] Online Appendix Figure K2 further shows that the effects on STEM enrolment are mainly driven by girls in the top quartile of maths performance.

[38] The average interval between the visits and the date when students completed the survey, is also comparable between the two groups of role models (Online Appendix Table E9, panel B).

[39] The significantly larger impact of professionals on grade 12 girls' probability of enrolling in STEM programs in general is also robust to controlling for these interactions (results available upon request).

[40] We also explored whether the effects of the interventions could be mediated by the subsequent interactions between the students and the teacher who was present during the visit. For instance, science teachers might be inclined to reiterate the role model's messages about science-related careers while female teachers might amplify the effects of the interventions for female students. Using data from the role model survey, we find no support for these hypotheses (results available upon request): the treatment effects on the STEM enrolment outcomes of girls in grade 12 do not vary significantly according to the teacher's gender or subject taught.

Table 8. *Heterogeneous Treatment Effects on Grade 12 Students' Outcomes, by Maths Performance.*

	Girls			Boys		
	Performance in maths			Performance in maths		
	Below median (1)	Above median (2)	p-value of diff. [q-value] (3)	Below median (4)	Above median (5)	p-value of diff. [q-value] (6)
Panel A. Enrolment outcomes						
Undergraduate major: STEM	0.007	0.028	0.567	−0.036	0.012	0.241
	(0.021)	(0.026)	[0.639]	(0.027)	(0.029)	[0.484]
Undergraduate major: selective STEM	−0.004	0.065***	0.011	−0.019	0.023	0.232
	(0.013)	(0.022)	[0.022]	(0.019)	(0.026)	[0.464]
Undergraduate major: male-dominated STEM	0.020	0.042*	0.506	−0.001	0.016	0.672
	(0.018)	(0.023)	[0.507]	(0.026)	(0.028)	[0.673]
N	1,544	1,211		1,497	1,328	
Panel B. Student perceptions						
Positive perceptions of science-related careers (index)	0.239***	0.340***	0.263	0.053	0.275***	0.011
	(0.053)	(0.058)	[0.440]	(0.056)	(0.052)	[0.100]
More men in science-related jobs	0.152***	0.078***	0.056	0.159***	0.143***	0.609
	(0.026)	(0.024)	[0.339]	(0.025)	(0.019)	[0.822]
Equal gender aptitude for maths (index)	0.034	0.134***	0.151	0.055	0.212***	0.083
	(0.043)	(0.045)	[0.339]	(0.063)	(0.062)	[0.375]
Women do not really like science	0.023	0.061***	0.130	0.077***	0.072***	0.874
	(0.015)	(0.016)	[0.339]	(0.025)	(0.022)	[0.874]
Women face discrimination in science-related jobs	0.096***	0.088***	0.845	0.097***	0.047	0.268
	(0.027)	(0.031)	[0.845]	(0.031)	(0.028)	[0.484]
Taste for science subjects (index)	−0.049	0.040	0.293	−0.007	0.031	0.658
	(0.051)	(0.056)	[0.440]	(0.063)	(0.053)	[0.822]
Self-concept in maths (index)	0.063	−0.045	0.145	0.092*	0.019	0.249
	(0.049)	(0.051)	[0.339]	(0.048)	(0.041)	[0.484]
Science-related career aspirations (index)	0.064	0.134**	0.397	0.042	0.071	0.730
	(0.049)	(0.060)	[0.511]	(0.056)	(0.053)	[0.822]
N	1,420	1,123		1,340	1,227	

Notes: This table reports estimates of the treatment effects of the role model interventions on grade 12 students' outcomes, separately by gender and performance in maths. Each row corresponds to a different linear regression performed separately by gender, with the dependent variable listed on the left. Students' performance in maths is measured from the grades obtained on the final maths exam of the *baccalauréat*. Columns (1) and (2) (for girls) and columns (4) and (5) (for boys) report the LATE estimates for students below and above the median level of performance in maths, respectively. They are obtained from a regression of the outcome of interest on the interaction between a classroom visit indicator and indicators for the student being below or above the median level of performance in maths, using treatment assignment (interacted with the maths performance dummies) as an instrument for treatment receipt. The regression includes school fixed effects (to account for the fact that randomisation was stratified by school) and the student characteristics listed in Table 1. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomisation (class). Columns (3) and (6) report both the cluster-robust model-based p-value for the difference between the treatment effect estimates for students above versus below the median performance in maths and, in square brackets, the p-value (q-value) adjusted for multiple hypothesis testing, using the FDR control method. Specifically, we use the sharpened two-stage q-values introduced in Benjamini *et al.* (2006) and described in Anderson (2008). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 9. *Treatment Effects (ITT) on Enrolment in a Selective STEM Program for Grade 12 Students: Heterogeneity by Student and Role Model Characteristics.*

	Dependent variable: enrolled in a selective STEM program					
	Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment group indicator (*T*) interacted with student characteristics						
$T \times$ baccalauréat percentile rank in maths (/100, demeaned)	0.138***	0.153***	0.155***	0.060	0.015	0.013
	(0.048)	(0.055)	(0.054)	(0.055)	(0.057)	(0.059)
$T \times$ baccalauréat percentile rank in French (/100, demeaned)		−0.036	−0.030		0.083	0.082
		(0.043)	(0.043)		(0.056)	(0.056)
$T \times$ high SES, demeaned		0.032	0.026		−0.013	−0.018
		(0.026)	(0.027)		(0.031)	(0.031)
Treatment group indicator (*T*) interacted with role model characteristics						
$T \times$ professional	0.059***	0.066***	0.102***	0.055**	0.048*	0.070*
	(0.020)	(0.020)	(0.025)	(0.027)	(0.028)	(0.036)
$T \times$ participated in the program the year before			−0.047**			0.024
			(0.023)			(0.042)
$T \times$ age (demeaned)			0.001			−0.000
			(0.001)			(0.002)
$T \times$ non-French			−0.022			−0.006
			(0.026)			(0.050)
$T \times$ has children			0.029			0.034
			(0.026)			(0.035)
$T \times$ has a PhD degree			0.091***			0.057
			(0.023)			(0.039)
$T \times$ field: maths, physics, engineering			−0.040*			−0.035
			(0.024)			(0.030)

Table 9. *Continued*

	Dependent variable: enrolled in a selective STEM program					
	Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
Other controls						
Treatment group indicator (T)	Yes	Yes	Yes	Yes	Yes	Yes
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
T interacted with school characteristics	No	Yes	Yes	No	Yes	Yes
T interacted with month of intervention	No	No	Yes	No	No	Yes
Observations	2,827	2,827	2,827	2,924	2,924	2,924
Adjusted R-squared	0.125	0.125	0.127	0.196	0.195	0.194

Notes: Each column corresponds to a separate regression. The sample is restricted to students in grade 12 (science track). The outcome variable is an indicator for being enrolled in a selective STEM undergraduate program in the year following high school graduation, i.e., 2016/17. The models are estimated separately for girls (columns (1)–(3)) and boys (columns (4)–(6)). The coefficients reported in columns (1) and (4) are from a regression of the outcome variable on a treatment group indicator (T), student characteristics, school fixed effects and the treatment group indicator interacted with the student's *baccalauréat* percentile rank in maths (between 0 and 1) and with an indicator for the role model being a professional. The specification in columns (2) and (5) includes further interactions between the treatment group indicator and both student and school characteristics. Finally, the specification in columns (3) and (6) adds interactions between the treatment group indicator and the characteristics of role models, as well as interactions between the treatment group indicator and dummies for the month of intervention. The student characteristics are those listed in Table 1, as well as the student's percentile ranks on the *baccalauréat* final exams in maths and French. The role model characteristics consist of age and a set of indicators for being a professional, having participated in the program the year before, being non-French, having children, holding a PhD degree and having graduated from a male-dominated STEM field (maths, physics, engineering). The school characteristics are dummies for the regional education authority where the high school is located (Paris, Créteil and Versailles) and a dummy for whether the school is private. School characteristics are only included through their interactions with the treatment group indicator, as these characteristics are absorbed by the school fixed effects. Since each high school was visited by at most one role model, role model and month-of-visit fixed effects are also absorbed by the school fixed effects. Standard errors (in parentheses) are adjusted for clustering at the class level. Observations with some missing characteristics are included in the regressions. An arbitrary value is assigned to all the missing characteristics and a set of dummy variables is created, with each variable being equal to one if the corresponding information is missing. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

6.2. *Heterogeneous Effects on Student Perceptions*

6.2.1. *Role model background: researchers versus professionals*

Why were the professionals more effective than the researchers in influencing female students' choices of study? To investigate this question, we examine how the two groups managed to change students' perceptions. We consider as potential channels of influence the dimensions studied in Section 4, namely general perceptions of science-related careers and gender roles in science, taste for science subjects, self-concept in maths and science-related career aspirations.

For girls in grade 12, a key finding is that professionals and researchers were equally successful in debunking stereotypes on gender differences in maths aptitude, and that they reinforced students' perceptions that 'women do not really like science' and that 'women face discrimination in science-related jobs' to a comparable extent (Table 7, panel B).[41] These results suggest that the 'gender debiasing' component of the classroom interventions, which emphasised men's and women's equal predisposition for science, cannot explain, alone, why the interventions increased girls' enrolment in selective STEM; otherwise, the two groups of role models would be expected to have had the same effect on enrolment outcomes, which is not what we find. By contrast, Table 7 reveals that in grade 12, the professionals improved female students' perceptions of science-related jobs more than the researchers and stimulated their aspirations for such careers more strongly. These dimensions thus seem more likely to explain why the professionals had a stronger influence on female students' choices of study.[42]

6.2.2. *High versus low achievers in maths*

The differences in the effects on students' perceptions are less pronounced between girls above and below the median of maths performance in grade 12 (Table 8, panel B). Although the differences are not statistically significant, it is interesting to note that high-achieving girls seem to have been more receptive to the messages that the professionals were better at conveying. Indeed, the point estimates suggest that perceptions of science-related careers improved more among the girls with above-median performance in maths. Aspirations for science-related careers also increased more among these girls, whereas awareness of female under-representation in science-related jobs increased less.

6.3. *A Generalisation Using Machine Learning Techniques*

Investigating treatment effect heterogeneity by splitting the sample into subgroups inevitably entails the risk of data mining. To address this concern, we carry out a systematic exploration of heterogeneous treatment effects using machine learning (ML) methods (see Athey and Imbens, 2017 for a review). Specifically, we adopt the approach developed by Chernozhukov *et al.* (2018) to estimate conditional average treatment effects. A detailed description can be found in Online Appendix L. Essentially, this approach allows us to compare the characteristics of the students

[41] The results for girls and boys in grade 10 are presented in Online Appendix Table K1. In this grade level, the effects of the two types of role model on girls' perceptions are more similar, and are not significantly different after adjusting for multiple hypothesis testing.

[42] The results for boys in grade 10 (Online Appendix Table K1, columns (4) to (6)) and grade 12 (Table 7, columns (4) to (6)) do not show substantial differences in effects. If anything, the professionals seem to have been slightly more effective than the researchers in increasing grade 10 boys' taste for science and debunking their stereotyped views on gender differences in maths aptitude.

whose educational choices were the most and the least affected by the classroom interventions. This first step serves to confirm, in a more agnostic way, the insights obtained from the comparison between professionals and researchers, and between high and low achievers in maths. Building on Chernozhukov *et al.* (2018), we then use a novel method to estimate the correlation between the treatment effects on enrolment outcomes and the effects on student perceptions. This second step takes advantage of the predicted heterogeneity in treatment effects by student and role model characteristics, to identify the messages that had the greatest impact on students' educational choices.

6.3.1. *Heterogeneous treatment effects on enrolment outcomes*

The results from the estimation of heterogeneous treatment effects on enrolment outcomes after grade 12 are reported in Online Appendix Tables L1 and L2, and are described in detail in Online Appendix L. The ML approach of Chernozhukov *et al.* (2018) confirms that there is considerable heterogeneity in treatment effects on selective STEM enrolment among girls in grade 12: they range from a small negative impact for the least affected quintile of girls to a large and significant 13.9 pp increase for the most affected quintile.[43] Consistent with the results discussed in Section 6.1, the comparison of the characteristics of the most and least affected quintiles confirms that role model background and student maths performance are the two main observable dimensions of heterogeneity: the average gap in maths performance rank between girls in the top and bottom quintiles of predicted treatment effects on selective STEM enrolment is as much as 63 percentiles, and the difference in the probability that the class was visited by a professional is 14.8 pps.

6.3.2. *Heterogeneous treatment effects on potential channels*

The results for heterogeneous treatment effects on student perceptions are reported in Online Appendix Table L3. For each possible channel, we compare the average maths performance of grade 12 girls in the top and bottom quintiles of predicted treatment effects, as well as their probability of being exposed to a professional rather than a researcher. The results confirm that the role models with a professional background conveyed a positive image of science, and raised girls' aspirations for science careers significantly more than the researchers. The ML approach also shows that the professionals were significantly less likely than the researchers to increase grade 12 girls' awareness of the under-representation of women in science-related jobs: compared to the least affected quintile of girls for this outcome, the most affected quintile is 11.2 pps more likely to have been visited by a researcher. These results are consistent with the notion that gender-neutral messages about careers in science are more effective than gender-related messages in steering girls towards STEM.

Regarding maths performance, the ML approach broadly confirms the insights from the subgroup comparisons presented in Section 6.2, but it appears better suited to reveal significant contrasts. Average maths performance is found to be significantly better among the girls whose perceptions of science-related careers and taste for science subjects improved the most. Conversely, maths performance is significantly poorer among those whose awareness of female under-representation in STEM and perception of gender discrimination increased the most.

[43] The lesser heterogeneity in the effects on enrolment in male-dominated STEM is also confirmed, with no statistically significant difference between the top and bottom quintiles of predicted treatment effects.

6.3.3. *Correlation between treatment effects*

So far, our discussion of the channels of influence has sought to identify the main dimensions of treatment effect heterogeneity on STEM enrolment outcomes, and has investigated how the effects on student perceptions vary along these dimensions. We now present results from a more general approach that builds on Chernozhukov *et al.* (2018) to produce a direct estimate of the correlation between the treatment effects on different outcomes conditional on exogenous observable characteristics. This approach, whose details are provided in Online Appendix L.3, constitutes a methodological contribution that can be used in other randomised controlled trials to relate treatment effects on different outcomes. In our context, the method allows us to determine whether, given their observable characteristics, the students with the largest treatment effects for a potential channel of influence Y^A are the same as those who exhibit the largest treatment effects on enrolment outcome Y^B.

We use this approach to estimate the correlation between the treatment effects for girls in grade 12 (see Online Appendix Table L5). The results confirm that some channels are more important than others in steering female students towards STEM studies. In particular, we find that the treatment effects on girls' enrolment in selective STEM exhibit a strong and significant positive correlation with the improvement in their perceptions of science-related careers ($\hat{\rho} = 0.96$), and a weaker positive correlation with their increased aspirations for such careers ($\hat{\rho} = 0.36$). By contrast, debiasing girls' attitudes towards gender differences in maths aptitude is not strongly associated with increased enrolment in selective STEM programs ($\hat{\rho} = 0.19$) and, if anything, reinforcing the belief that women suffer discrimination in science careers tends to deter girls from enrolling in these programs ($\hat{\rho} = -0.34$).

Overall, the results based on the correlations between treatment effects are consistent with and extend those obtained earlier. They suggest that the most effective role models were those who managed to convey a positive image of science careers, and raise students' aspirations without stressing women's under-representation and its possible causes too strongly. These features are in line with the main mechanisms usually considered necessary for role models to work: generating a sense of fit while moderating the effects of stereotype threat.

7. Conclusion

Based on a large-scale randomised field experiment involving 56 female role models and nearly 20,000 grade 10 and grade 12 students, this paper shows that a one-hour in-class exposure to a woman scientist can improve students' perceptions of science careers, and significantly increase female participation in STEM fields of study at college enrolment. Remarkably, the positive enrolment effects are observed only in the academic tracks with the most severe gender imbalance, which are the most prestigious and selective, and those that are most maths intensive. These effects can be expected to increase the future earnings of the target population, since the selective and male-dominated STEM programs offer large wage premiums relative to other programs.

We analyse the channels that could explain these significant effects on enrolment outcomes. We show that the classroom interventions had no discernable effect on students' academic performance and improved their self-concept in maths only slightly, ruling these factors out as primary causes. By contrast, the visits significantly challenged students' stereotyped views of science careers and gender differences in aptitude for science. These effects, however, are observed for both genders in both grades, suggesting that by themselves they cannot explain why the role model interventions affected only the educational choices of girls in grade 12. Rather, our results offer substantial evidence that female students' responses to the role model interventions were mediated by their ability to identify with the female scientists to whom they were exposed. Girls in grade 12 were more receptive than the other groups of students to the appealing image of science-related careers embodied by the role models, and their aspirations for such careers increased substantially. This process of identification was less likely to occur among grade 10 girls, who are further away from career choices, and for boys at both grade levels, who may have found it more difficult to identify with women scientists.

A central finding is that the effects on grade 12 girls' educational choices varied markedly with the scientists who conducted the classroom visits. This significant heterogeneity demonstrates that role model interventions are not reducible to information provision, and highlights the importance of the role models' profile in generating a sense of fit among students. In our experiment, women with a professional background were more effective than researchers in conveying an attractive image of careers in science and elevating girls' aspirations. Our results thus suggest that these are critical skills to target when choosing role models (teachers, instructors, career women, etc.). While our ability to pinpoint the attributes that were to the advantage of the professionals in our setting is limited, a likely explanation is that they were better established in their careers and had better paying jobs than the researchers. The role played by these channels would be a fruitful topic for future research.

Another important insight from the study is that, by heightening awareness of the under-representation of women in STEM, while at the same time observing that men and women have equal aptitude for science, the interventions may have unintentionally reinforced students' beliefs that women dislike science and face discrimination in STEM careers. That is, there is suggestive evidence that overemphasising gender can be counter-productive and that gender-neutral messages might be more effective in steering girls towards STEM fields. In our setting, the role models who reinforced the perception that women are under-represented and discriminated against in science had the least effect on selective STEM enrolment for female students in grade 12, whereas those who improved girls' perceptions of science careers the most had the greatest impact. These findings suggest that role model interventions need to be carefully designed to limit the potential discouragement effect of overemphasis on gender imbalances.

More generally, our heterogeneity analysis warns against the temptation to view role models as a one-size-fits-all remedy for women's under-representation in STEM fields. We find that the role model effects on enrolment outcomes are concentrated among high-achieving girls in maths. The effectiveness of this type of intervention in increasing female participation in STEM among lower-performing students remains an open question that warrants further research.

CNRS & Paris School of Economics, France
CNRS & Paris School of Economics, France
IREDU-Université de Bourgogne, France
University of Toronto, Canada

Additional Supporting Information may be found in the online version of this article:

Online Appendix
Replication Package

References

Anderson, M.L. (2008). ‘Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects’, *Journal of the American Statistical Association*, vol. 103(484), pp. 1481–95.

Athey, S. and Imbens, G.W. (2017). ‘The econometrics of randomized experiments’, in (E. Duflo and A.V. Banerjee, eds.), *Handbook of Economic Field Experiments*, vol. 1, pp. 73–140, Amsterdam: Elsevier.

Avvisati, F., Gurgand, M., Guyon, N. and Maurin, É. (2014). ‘Getting parents involved: A field experiment in deprived schools’, *Review of Economic Studies*, vol. 81(1), pp. 57–83.

Beaman, L., Duflo, E., Pande, R. and Topalova, P. (2012). ‘Female leadership raises aspirations and educational attainment for girls: A policy experiment in India’, *Science*, vol. 335(6068), pp. 582–86.

Beede, D., Julian, T., Langdon, D., McKittrick, G., Khan, B. and Doms, M. (2011). ‘Women in STEM: A gender gap to innovation’, Issue Brief 04-11, U.S. Department of Commerce, Economics and Statistics Administration.

Benjamini, Y., Krieger, A.M. and Yekutieli, D. (2006). ‘Adaptive linear step-up procedures that control the false discovery rate’, *Biometrika*, vol. 93(3), pp. 491–507.

Bettinger, E.P. and Long, B.T. (2005). ‘Do faculty serve as role models? The impact of instructor gender on female students’, *American Economic Review*, vol. 95(2), pp. 152–7.

Black, D.A., Haviland, A.M., Sanders, S.G. and Taylor, L.J. (2008). ‘Gender wage disparities among the highly educated’, *Journal of Human Resources*, vol. 43(3), pp. 630–50.

Blau, F.D. and Kahn, L.M. (2017). ‘The gender wage gap: Extent, trends, and explanations’, *Journal of Economic Literature*, vol. 55(3), pp. 789–865.

Breda, T., Grenet, J., Monnet, M. and Van Effenterre, C. (2016a). ‘Evaluation of the For Girls in Science intervention—role model survey [database]: rm_survey’, https://doi.org/10.5281/zenodo.7588802.

Breda, T., Grenet, J., Monnet, M. and Van Effenterre, C. (2016b). ‘Evaluation of the For Girls in Science intervention—role model characteristics [database]: rm_info’, https://doi.org/10.5281/zenodo.7588802.

Breda, T., Grenet, J., Monnet, M. and Van Effenterre, C. (2016c). ‘Evaluation of the For Girls in Science intervention—student survey [database]: stu_survey’, https://doi.org/10.5281/zenodo.7588802.

Brown, C. and Corcoran, M. (1997). ‘Sex-based differences in school content and the male-female wage gap’, *Journal of Labor Economics*, vol. 15(3), pp. 431–65.

Bursztyn, L. and Yang, D.Y. (2022). ‘Misperceptions about others’, *Annual Review of Economics*, vol. 14(1), pp. 425–52.

Canaan, S. and Mouganie, P. (2021). ‘The impact of advisor gender on female students’ STEM enrollment and persistence’, *Journal of Human Resources*, 0320–10796R2.

Carlana, M. (2019). ‘Implicit stereotypes: Evidence from teachers’ gender bias’, *Quarterly Journal of Economics*, vol. 134(3), pp. 1163–224.

Carrell, S.E., Page, M.E. and West, J.E. (2010). ‘Sex and science: How professor gender perpetuates the gender gap’, *Quarterly Journal of Economics*, vol. 125(3), pp. 1101–44.

CGE. (2018). ‘L’insertion des diplômés des grandes écoles. Résultats de l’enquête 2018’, Conférence des Grandes Écoles (CGE), Paris, https://www.cge.asso.fr/themencode-pdf-viewer/?file=https://www.cge.asso.fr/wp-content/uploads/2018/06/2018-06-19-Rapport-2018.pdf. (last accessed: 28 August 2019).

Chernozhukov, V., Demirer, M., Duflo, E. and Fernández-Val, I. (2018). ‘Generic machine learning inference on heterogenous treatment effects in randomized experiments’, Working Paper 24678, National Bureau of Economic Research.

Coffman, K.B. (2014). ‘Evidence on self-stereotyping and the contribution of ideas’, *Quarterly Journal of Economics*, vol. 129(4), pp. 1625–60.

DAPEP. (2017). ‘Base Élèves Académique (BEA) [database]: ELH 2012–2014, ELC 2013–2016, ELG 2015’, Délégation Académique à la Prospective et à l’Évaluation des Performances (DAPEP), Rectorat de l’Académie de Versailles.

Del Carpio, L. and Guadalupe, M. (2021). ‘More women in tech? Evidence from a field experiment addressing social identity’, *Management Science*, vol. 68(5), pp. 3196–218.

Goldin, C. (2014). ‘A grand gender convergence: Its last chapter’, *American Economic Review*, vol. 104(4), pp. 1091–119.

Hoogendoorn, S., Oosterbeek, H. and van Praag, M. (2013). ‘The impact of gender diversity on the performance of business teams: Evidence from a field experiment’, *Management Science,*, vol. 59(7), pp. 1514–28.

Lavy, V. and Sand, E. (2018). ‘On the origins of gender gaps in human capital: Short- and long-term consequences of teachers’ biases’, *Journal of Public Economics*, vol. 167(C), pp. 263–69.

Lemaire, S. (2012). ‘Les bacheliers S: Des poursuites d’études de plus en plus dispersées’, Note d’Information 12.10, MESR-SIES.

Lim, J. and Meer, J. (2017). ‘The impact of teacher-student gender matches: Random assignment evidence from South Korea’, *Journal of Human Resources*, vol. 52(4), pp. 979–97.

MENJ-DEPP. (2017). ‘Organisation des Concours et Examens Nationaux (OCEAN) [database]: OCEAN-DNB 2010–2015, OCEAN-BAC 2015 and 2016’, Ministère de l’Éducation Nationale et de la Jeunesse.

MESRI-DGESIP/DGRI-SIES. (2017). ‘Système d’Information sur le Suivi de l’Étudiant (SISE) [database]: SISE-UNIV 2013–2016, SISE-ENS 2013–2016, SISE-INGE 2013–2016, SISE-MANA 2013–2016, SISE-PRIV 2013–2016’, Ministère de l’Enseignement supérieur, de la Recherche et de l’Innovation (MESRI), Direction générale de l’enseignement supérieur et de l’insertion professionnelle (DGESIP), Direction générale de la recherche et de l’innovation (DGRI).

MESRI. (2018). 'Enquête d'Insertion Professionnelle à 30 Mois des Diplômés de Master 2015 [database]', Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI), https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-insertion_professionnelle-master_donnees_nationales/information/ (last accessed: 28 August 2019).

National Science Foundation. (2017). 'Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017', Special Report NSF 17-310, National Science Foundation and National Center for Science and Engineering Statistics.

Nguyen, T. (2008). 'Information, role models and perceived returns to education: Experimental evidence from Madagascar', Unpublished manuscript, Massachusetts Institute of Technology.

PAPP. (2017). 'Base Élèves Académique (BEA) [database]: ELH 2012–2014, ELC 2013–2016, ELG 2015', Pôle Académique de la Prospective et de la Performance (PAPP), Rectorat de l'Académie de Créteil.

Porter, C. and Serra, D. (2020). 'Gender differences in the choice of major: The importance of female role models', *American Economic Journal: Applied Economics*, vol. 12(3), pp. 226–54.

Riise, J., Willage, B. and Alexander, W. (2022). 'Can female doctors cure the gender STEMM gap? Evidence from randomly assigned general practitioners', *The Review of Economics and Statistics*, vol. 104(4), pp. 621–35.

SSA. (2017). 'Base Élèves Académique (BEA) [database]: ELH 2012–2014, ELC 2013–2016, ELG 2015', Service Statistique de l'Académie de Paris (SSA), Rectorat de l'Académie de Paris.

Terrier, C. (2020). 'Boys lag behind: How teachers' gender biases affect student achievement', *Economics of Education Review*, vol. 77, 101981.

Weinberger, C.J. (1999). 'Mathematical college majors and the gender gap in wages', *Industrial Relations*, vol. 38(3), pp. 407–13.