Can Female Role Models Reduce the Gender Gap in Science? Evidence from Classroom Interventions in French High Schools*

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

This version: December 2018
(Preliminary – Please do not cite or quote without permission)

Abstract

This paper reports the results of a large-scale field experiment that was designed to assess whether short classroom interventions by external female role models with a science background can influence students' attitudes towards science-related careers and affect their choice of field of study. Using a random assignment of the interventions to high school classrooms in the Paris Region, we find that a one-hour exposure to a female role model increases by respectively 30 percent (20 percent) the probability for girls in Grade 12 to enroll in a selective (male-dominated) STEM track in higher education the following year, inducing an increase in the representation of girls in those tracks from 30 to 34 percent (28 to 31 percent). We find only limited effects of the interventions on boys' educational choices in Grade 12, and no effect for female and male students in Grade 10. Several mechanisms can explain the observed changes in college major choices among girls in Grade 12. First, for all students, the program strongly reduces the prevalence of stereotypes associated with jobs in science and gender roles in science. Second, it raises students' interest in science-related careers. Third, it slightly improves their math self-concept. We find that the program was particularly effective at steering high-achieving girls in Grade 12 towards selective STEM studies, and that female facilitators with a professional background had larger effects than young researchers. The results suggest, however, that role models also increase students' awareness of female underrepresentation in science and reinforce the belief that women are discriminated in STEM careers, which could explain their limited effects among low-achieving girls.

JEL codes: C93, I24, J16
Keywords: *Role Models*; *Gender*; *STEM*; *Stereotypes*; *Track Choice.*

*Breda: CNRS, Paris School of Economics, 48 boulevard Jourdan, 75014, Paris, France (e-mail: thomas.breda@ens.fr); Grenet: CNRS, Paris School of Economics, 48 boulevard Jourdan, 75014, Paris, France (e-mail: julien.grenet@psemail.eu); Monnet: Paris School of Economics, 48 boulevard Jourdan, 75014, Paris, France (e-mail: marion.monnet@ipp.eu); Van Effenterre: Harvard Kennedy School, 79 JFK Street, Cambridge, MA 02138, USA (e-mail: clementine_van_effenterre@hks.harvard.edu). We are grateful to 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 Educational authorities (Rectorats) of Créteil, Paris and Versailles for their invaluable assistance in collecting the data. This paper greatly benefited from discussions and helpful comments from Esther Duflo, Lena Edlund, Ruth Fortmann, Marc Gurgand, Élise Huillery, Sandra McNally, and Amanda Pallais. We thank participants in the 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 in Harvard, and Journées LAGV 2018 in Aix-Marseille. We also thank seminar participants at LSE, Harvard Kennedy School, Maastrict SBE, and PSE. We are grateful to the Institut des politiques publiques (IPP) for continuous support. The project received IRB approval at J-PAL Europe and was registered in the AEA RCT Registry with ID AEARCTR-0000903.

Introduction

Women's increasing level of participation in science and engineering fields in the U.S. has leveled off in the past decade (National Science Foundation, 2017). Over the period 2004–2014, the share of bachelor's degrees awarded to women in engineering and computer science has stagnated around 20 percent, while it has decreased from 46 percent to 43 percent in mathematics and statistics and from 42 percent to 40 percent in physical science.[1] These trends, which are observed in almost all OECD countries (OECD, 2016), are a source of concern for two main reasons.[2] First, they exacerbate gender inequality in the labor market, as Science, Technology, Engineering, and Mathematics (STEM) occupations offer better wages (Brown and Corcoran, 1997; Black et al., 2008; Blau and Kahn, 2017). Second, these trends lead to an increasing loss of talent that could reduce aggregate productivity (Weinberger, 1999) and they contribute to the shortage of STEM workers to meet the demands of the labor market in advanced economies (Carnevale et al., 2011).[3]

The origins of the gender gap in enrollment in STEM have been the focus of extensive research. Gender differences in ability or test scores can at most explain only a small fraction of this gap (Hyde, 2005; Spelke, 2005; Card and Payne, 2017) and some scholars argue that direct discrimination is no longer a major determinant (Ceci et al., 2014). Attention has increasingly been devoted to better understand gender differences in the educational choices made by equally able students. Parents, schools, and teachers are often said to convey stereotypes and social norms that influence these choices and contribute to maintain a strong gender segregation across school majors in the long run.[4]

One approach that has been put forward to limit these early influences is to expose students to successful or admirable people who do not embody the social norm. Such

[1] See `https://www.nsf.gov/statistics/2017/nsf17310/data.cfm`.

[2] On average across OECD countries, less than one in three engineering graduates and less than one in five computer science graduates are women. See `http://www.oecd.org/gender/data/shareofwomengraduatesbyfieldofeducation.htm`.

[3] Some economic projections advance that over the next decade, the U.S. would need approximately one million more STEM professionals than the country will produce at the current rate in order to retain its historical preeminence in science and technology (Olson and Riordan, 2012).

[4] These social pressures and gender stereotypes might not necessarily translate into explicit discrimination (Ceci and Williams, 2011; Breda and Ly, 2015; Breda and Hillion, 2016) but rather seem to be mostly internalized and thereby influence academic self-perception (Correll, 2001; Ehrlinger and Dunning, 2003), behavior in competitive environments (Gneezy et al., 2003; Niederle and Vesterlund, 2007, 2010; Buser et al., 2014), and foster an environment conducive to self-censorship (Babcock and Laschever, 2012; Leibbrandt and List, 2015).

"role models" may extend students' possibility sets, raise aspiration, alleviate stereotype threats, and provide relevant information. In that vein, a large body of work has established that female science teachers and professors may serve as role models and that they contribute to improve women's achievement (e.g., Dee, 2007; Hoffmann and Oreopoulos, 2009; Lim and Meer, 2017; Eble and Hu, 2017) and enrollment in STEM majors (Bettinger and Long, 2005; Carrell et al., 2010).[5] However, the policy implications of these are unclear. Substantially increasing the share of female students exposed to female instructors is not feasible in the short run, as the latter remain a scarce resource. Moreover, it is not possible to know for sure if the observed changes in educational choices are actually due to female professors acting as role models or if they are rather triggered by gender differences in teaching practices and behavior in class. For example, the positive influence of female instructors on female students might stem more from the stronger encouragements received over an entire academic year (Terrier, 2016; Lavy and Sand, 2018) than from self-identification with a successful example that does not embody the social norm. Carlana (2018) provides yet another mechanism likely to challenge explanations solely relying on role models. The author shows that implicit biases, as measured by the Gender-Science Implicit Association Test, are more pronounced among male teachers than among their female colleagues, and that these biases affect students' educational choices. Such biases could therefore explain why female students exposed to female science instructors are more likely on average to enroll in a science major.

In this paper, we overcome some of the above-mentioned limitations by testing, based on a large-scale randomized experiment, whether a one-hour exposure to external (non-teaching) female scientists acting as role models can change high school girls' perceptions of science careers and ultimately steer them towards STEM studies. Such short interventions are policy relevant as they can easily be scaled up, and they are more likely to capture the pure effect of role models.

We evaluate an awareness program that was launched in 2014 by the corporate foundation of a large French company with the aim of encouraging girls to explore STEM career paths. The program consists in a one-hour intervention by either a young female scientist (Ph.D. student or postdoc) or a young professional with a science background.

[5] Seminal papers on the impact of teacher-student gender interactions include Canes and Rosen (1995), Rothstein (1995), and Neumark and Gardecki (1998). More recent papers have also investigated teacher-student interactions in developing countries (Paredes, 2014; Muralidharan and Sheth, 2016).

These facilitators provide information on science-related careers in general and more specifically on the underrepresentation of women in science based on a set of customizable slides. They are also strongly encouraged to talk about their own experience and career path. The evaluation took place during the academic year 2015-2016 in 98 of the 489 private and public high schools located in the Paris region. It concerned approximately 14,000 students in close to 400 Grade 10 classes (*Seconde générale et technologique*)—at the end of which irreversible track choices are made—and approximately 6,000 students in about 200 Grade 12 science track classes (*Terminale scientifique*), which is the final year of secondary education in France. About half of these classes were randomly selected to be visited by one of the 56 facilitators who participated in the evaluation.

The present study contributes to a very narrow but growing literature that has documented positive effects of female role models in other contexts. Beaman et al. (2012) find that exposure to women in leadership positions in India has a positive impact on girls' educational attainment and parents' career aspirations for their daughter (but not for their son). More closely related to our paper is a recent study by Porter and Serra (2017), which documents a positive impact of two female role models who were carefully selected among the economics alumni of the Southern Methodist University in the U.S. on the likelihood that female students in that university enroll in an economics major. In contrast to these studies, we focus on the effect of external role models on the gender gap in STEM.

Building on comprehensive administrative data on students in secondary and higher education, we show that the program had no effect on Grade 10 students' educational choices, but increased by approximately 10 percent the probability that girls in Grade 12 enroll in STEM majors after graduating from high school. The effects are driven by the selective STEM tracks, which in France lead to the most prestigious graduate schools, and the most male-dominated ones (math, physics, computer sciences, engineering). The probability for girls in Grade 12 to choose these tracks increases by respectively 30 percent and 20 percent due to the program. Small positive effects are also found for boys, but only in the most prestigious selective STEM tracks. Combining the effects for girls and boys, we find that a one-hour intervention by female role models participating in the program increased female representation from 30 to 34 percent in STEM selective tracks, and from 28 to 31 percent in male-dominated tracks.

To shed light on the mechanisms underlying these effects, we conducted a post-treatment student survey consisting of an eight-page questionnaire which was administered in class one to six months after the classroom interventions. We find that the treatment (i) significantly improved boys' and girls' perceptions of scientific jobs and careers and their ability to consider these jobs as potential career choices; (ii) had no significant effect on students' self-reported taste for science subjects; (iii) slightly increased students' math self-concept; and (iv) raised awareness regarding the underrepresentation of women in science and its potential causes. Regarding this last dimension, the treatment increased students' perception that women are underrepresented in science despite having the same aptitude for science studies and careers than men. This outcome of the program is not surprising in light of the key messages conveyed by the female facilitators. However, by making the underrepresentation of women in science more salient, the interventions also appear to have triggered an unintended *ex post* rationalization of this message: girls and boys in the treatment group are more likely to think that women are discriminated in science and that they do not like science. These indirect effects could have been counterproductive, by steering some female students away from STEM studies.

We perform simple comparisons across students and role models to comfort this intuition and study in a more systematic way the messages that were the most effective at switching students' behavior. In doing so, we attempt to identify, among the dimensions that were affected by the program (perceptions, self-concept, etc.), those that were the most conducive to a change in actual behavior (choosing STEM studies). This analysis contributes to the understanding of the mechanisms by which role models may affect choices, suggesting for example that the provision of non gender-neutral information may be counterproductive,[6] while identification with role models can improve self-concept.[7]

The remainder of the paper is organized as follows. Section 1 provides some institutional background on the French educational system and on the gender gap in STEM fields.

[6]In contrast, positive gender-neutral information seems to have contributed to the program's impact on girls' educational choices. This is in line with the results of Nguyen (2008) who evaluates three interventions in primary schools in rural Madagascar, which were designed to increase the perceived returns to education, through statistical information, role models, or both. Both programs providing statistical information had a positive impact on school attendance, performance on tests, future school enrollment, and total educational attainment, while the program with role models but no formal provision of information had positive but only small average effects.

[7]This analysis complements several lab experiments that have investigated how the appearance and behavior of female STEM role models affects their effectiveness (Lockwood and Kunda, 1997; Dasgupta and Asgari, 2004; Cheryan et al., 2011; Betz and Sekaquaptewa, 2012; O'Brien et al., 2016).

Section 2 describes the program and the experimental design. Section 3 presents the data and the empirical strategy. Section 4 analyzes the effects of the classroom interventions on student perceptions, self-concept and educational choices. Section 5 extends the analysis to the persistence of effects, the role of the timing of interventions, and potential spillovers. Mechanisms are presented in Section 6 and Section 7 concludes.

1 Institutional Background

Structure of the French educational system. In France, education is compulsory from the age of 6 to the age of 16, with the school year running from September to June. The school system consists of five years of elementary education (from Grade 1 to Grade 5) and eight years of secondary education, which in turn is divided into four years of middle school (*collège*, from Grade 6 to Grade 9) and three of high school (*lycée*, from Grade 10 to Grade 12). High school terminates with the *Baccalauréat* national exam, which is normally taken the year students turn 17 and is required to access higher education.

During secondary schooling, the tracking of students occurs at two critical stages (see Figure 1 in the Appendix). At the end of middle school, about two thirds of students are admitted to general and technological upper secondary education (*Seconde générale et technologique*), while the remaining third is tracked into vocational secondary education (*Seconde professionnelle*). After the first year of high school (Grade 10), the general and technological track is further split into two separate tracks: approximately 80 percent of the student are allowed to prepare a general *Baccalauréat*, while the remaining 20 percent are directed towards a technological *Baccalauréat* during the last two years of high school (Grades 11 and 12). Students preparing a general *Baccalauréat* are required to choose between three tracks when entering Grade 11, specializing in either science (*Première S*), humanities (*Première L*), or social sciences (*Première ES*). This is an important choice given that the curriculum and high school examinations are specific to each *Baccalauréat* track, and therefore directly affect students' educational opportunities and career prospects after high school graduation. It is for instance almost impossible for a student to be admitted to engineering or medical undergraduate programs without holding a *Baccalauréat* in Science. Students directed to the technological track are also required to choose between a set of STEM and non-STEM subtracks, which can restrict their possible choices of study in higher education.

College admissions. Starting in the Spring term, students in Grade 12 apply for admission to higher education programs through an online centralized admission platform called *Admission post-bac* (APB). Applicants are requested to select and to rank up to 36 programs by order of preference and can modify their rank-ordered lists until the end of May.[8] The programs to which students can apply fall into two broad categories, each accounting for approximately half of first-year undergraduate enrollment: (i) non-selective undergraduate university programs (*Licence*), which are open to all students who hold the *Baccalauréat*; and (ii) selective programs, which mainly include *Classes préparatoires aux grandes écoles* (CPGE), *Section de technicien supérieur* (STS) and *Institut universitaire de technologie* (IUT). Non-selective programs cannot select students based on their academic background: when the number of applications exceeds the number of seats available, students from the university's academic area (*académie*) are given priority over other students and, within the group of local applicants, students who have ranked the program at a higher position in their list are given priority; remaining ties are broken using a random lottery. By contrast, selective programs are free to select students based on their own criteria, including students' grades in the last two year of high schools, but without knowing at which position they are ranked by applicants. Among selective programs, the most prestigious are the two-year CPGE programs, which prepare students to take the national entry exams to elite graduate schools (*Grandes Écoles*). These programs are either specialized in science, economics and business, or humanities.[9] Taking as inputs program students' rank-ordered list of choices, programs' priority ranking of applicants and their capacities, the student-program matching is obtained after executing the college-proposing version of Gale and Shapley (1962)'s deferred acceptance mechanism.

Female underrepresentation in STEM fields of study. The proportion of female students in STEM-oriented studies drops gradually during high school and at entry in higher education (see Figure A1 in the Appendix). While 54 percent of students enrolled

[8]Students can rank up to 12 choices per program type (university, selective programs, two-year college/vocational training, art schools, schools of architecture, business schools, and engineering schools) and at most 36 choices in total.

[9]Within the scientific CPGE programs, students can choose between pure mathematics and physics programs (MPSI), physics and chemistry (PCSI), and biology/geoscience (BCPST).

in Grade 10 (academic track) are female, this share drops to 47 percent in the *Baccalauréat* science track (Grade 11 and Grade 12). Female representation in selective STEM undegraduate programs further shrinks to approximately 30 percent—this share having remained fairly stable over the past fifteen years (see Appendix Figure A2). Most female students who graduate from the Science track in high school opt for medical undergraduate studies, in which where they are over-represented (64 percent of enrollment).

2 Program and Experimental Design

2.1 The Program

The program being evaluated is an awareness campaign that was first launched in 2014 by the corporate foundation of a large French company, with the aim of encouraging girls to explore STEM career paths based on in-class interventions by female role models with a background in science. The program, which is targeted at female and male high school students, was first launched in France and was later expanded to other countries, including Canada, Italy and New Zealand. The class-based interventions are carried out by female facilitators with two types of background: (i) young professionals privately employed by the firm who volunteer for the program; and (ii) Ph.D. students or post-doctoral researchers who receive a research fellowship from the firm's foundation and who participate in the program as part of their contract.

Content of the interventions. The classroom interventions last one hour and are subdivided into four main sequences. The presentation begins with a set of customizable slides aimed at highlighting two key facts: (1) jobs in STEM are in high demand but face a severe shortage of graduates in the relevant fields of studies; and (2) women are underrepresented in STEM careers. These two facts are illustrated with examples of career prospects in humanities versus science, emphasizing differences in employment rates, average earnings, and the prevalence of gender segregation in high-wage occupations. The slides further stress the differences within science between STEM and non-STEM fields, and the contribution of female underrepresentation in STEM to the overall gender pay gap.

The second sequence kicks off with the screening of two three-minute videos aimed at illustrating and deconstructing stereotypical views about science-related careers and gender roles in science.[10] The first video, entitled "Science, Beliefs or Reality?," uses interviews with high school students to debunk myths about careers in science (e.g., occupations in science are more challenging, they all require long studies), stereotypes attached to scientists (e.g., they are introverted, lonely), and gender differences in intrinsic aptitude for science (e.g., women are less talented in math). The second video, entitled "Are we all Equal in Science?," describes the gender stereotypes usually attached to women in science while providing information on brain plasticity and on how interactions and the social environment shape men and women's abilities and tastes. This sequence is aimed at stimulating class discussion based on students' reactions to the videos.

The third sequence is centered around the female facilitator's own experience as a woman with a background in science and consists of an interactive question-and-answer session with the students. Examples of topics addressed during this discussion include the facilitator's typical day at work, her everyday interactions with co-workers, how much she earns, as well as work-family balance.

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

2.2 Experimental Design

Selection of schools and classes. The experiment was carried out in the three educational districts (*académies*) of the Paris region (Paris, Créteil and Versailles) during the 2015-2016 academic year, i.e., in the program's second year of existence. Together, the three districts include 736,000 high school students, representing 19 percent of total high school enrollment in France.[11] The districts of Créteil and Versailles are the two largest in France.

In the Spring of 2015, the French Ministry for Education agreed to support a randomized evaluation of the program. The Ministry designated three representatives (one in each educational district) to serve as intermediaries between the selected high schools and

[10]Thumbnails of the videos shown during the classroom interventions are dispalyed in Appendix Figure B4.

[11]Student shares are respectively 9.1 percent for Versailles, 7.1 percent for Créteil and 2.8 percent for Paris (Ministère de l'Éducation Nationale, 2014).

the evaluation team. In June 2015, official letters were sent out to inform all high school principals that they were likely to be contacted by the team of researchers to participate in the experiment. Between September and November 2015, all general high schools with at least four classes in Grade 10 and two classes in Grade 12 (science track) were contacted by our team, representing approximately 300 out of the 489 high schools operating in the three districts. Of these schools, 98 accepted to take part in the experiment, representing 10 percent of Grade 10 enrollment and 14 percent of Grade 12 enrollment in the three districts combined.[12] The overall experimental sample, which includes 19,451 students (13,700 in Grade 10 and 5,751 in Grade 12), is reasonably representative of the population of students in Grade 10 and Grade 12 (science track) in the Parisian region both in terms of its social composition and of average performance on the DNB exam—a national exam which students take at the end of middle school (see Table B1 in the Appendix).

Randomization. In the Fall of the academic year 2015-2016, the principals of participating high schools were invited to select at least six classes—four in Grade 10 and two in Grade 12 (science track)—and to indicate a preferred time slot and day of the week for the interventions.[13] The classes selected by the principals were subject to random assignment within each school, half of the classes—or half rounded up or down to the nearest integer in the rare cases where there was an odd number of selected classes (5 schools for Grade 10 classes and one school for Grade 12 classes)—being assigned to the treatment group. In total, 302 classes were randomly assigned to the treatment group to receive the visit of a female facilitator, the other 299 classes serving as a the control group. Table 1 indicates that the random assignment successfully balanced the characteristics of students in the treatment and control groups in the experimental sample.

Female facilitators. The experiment involved a total of 56 female facilitators, among which 35 were employees of the firm whose foundation is sponsoring the program and 21 were Ph.D. students or post-doctoral researchers receiving a research fellowship from the firm's foundation. Table 2 provides summary statistics of the facilitators' characteristics. Facilitators with a research background tend to be younger than the professionals

[12] The location of the high schools that participated in the experiment is displayed in Appendix Figure B5.

[13] In the vast majority of schools, principals selected exactly four Grade 10 classes and two Grade 12 classes. A small number of schools selected only Grade 10 classes due to timetable constraints for Grade 12 classes.

employed by the firms (30 vs. 36 years of age on average) and are less often of foreign nationality (10 percent vs. 17 percent). Although both types of facilitators have very high levels of educational attainment, with 40 percent having attended a Grande École, the researchers are more likely to hold a Ph.D. (95 percent vs. 39 percent) and to have graduated with a degree in math, physics and engineering (38 percent vs. 14 percent) than the facilitators with a professional background. They are also less likely to have children (19 percent vs. 56 percent) and to have been involved in the program in the previous year (19 percent vs. 29 percent). On average, each facilitator carried out five classroom interventions in two different high schools.

Classroom interventions. The classroom visits took place between November 17, 2015, and March 3, 2016.[14] Each facilitator was asked to select two to three schools in which to carry out an average of three classroom visits per schools—in most cases, two visits in Grade 10 and one visit in Grade 12. Facilitators were not randomly assigned to schools but rather selected the schools and time slots using an online system on a first-come, first-served basis.[15]

3 Data and Empirical Strategy

3.1 Data

To evaluate the program's effects on students' perceptions and educational choices, we combine three main data sources: (i) a post-visit survey of facilitators; (ii) a post-visit survey of students in the treatment and control classes; and (iii) individual-level administrative data on students.

Facilitator survey. After each visit to a school, which typically involved three consecutive classroom interventions, we asked the facilitators to complete an online survey to

[14] 16 percent of the classrooms were visited in November, 25 percent in December, 39 percent in January, 19 percent in February, and 1 percent in March.

[15] Randomly assigning the facilitators to the schools was not possible since most of them participated in the program on a voluntary basis and outside of their regular working hours. The non-random assignment of facilitators to schools does not, however, threaten our identification hypothesis since the random assignment of classes to the treatment and control groups was stratified by school. Moreover, as participating schools were only gradually added to the schedule, multiple registration sessions were organized to match the facilitators with the participating schools. All facilitators were contacted four times in total, on October 21, November 24, December 7, 2015, and on February 3, 2016.

collect general feedback. This survey also served the purpose of monitoring compliance with random assignment by asking the facilitators to report the name of each class they visited. The main summary statistics for the facilitator survey are shown in Appendix Table E5. The interventions almost always took place in the presence of a teacher (89 percent) and sometimes with another adult (35 percent). The facilitators reported organizational problems for only 16 percent of the visits (e.g., the intervention started late, the slides could not be shown). When asked about their overall perception of the intervention, 90 percent of the facilitators declared that the intervention went "well" (55 percent) or "very well" (35 percent). Students were generally perceived by the facilitators to be responsive to the key messages conveyed, especially those contained in the two short videos shown in class.

Student survey. We conducted a post-treatment survey in all treatment and control classes one to six months after the visits, between January and March 2016. Each paper-and-paper questionnaire was assigned a unique identifier so that it could be linked with individual-level administrative data. The survey was designed to collect a rich set of information on students' tastes, personality traits, choices and stereotypes, and was administered in exam conditions under the supervision of a teacher.[16]

The first part of the questionnaire asked students about their extracurricular activities (e.g., competitive sports, video games, etc.), how they assess their performance in different subjects taught at school, and whether they enjoy these subjects. We included two questions from the Programme of International Student Assessment (PISA) survey to measure students' self-concept and self-confidence in math, which have been shown to generate large gender gaps ("I am worried when I think about math" and "I am lost in front of a math problem").

The second and third parts of the questionnaire collect detailed information on students' intended field of study and career intentions, as well as on their attitudes towards science. We asked students whether they enjoy science in general, whether they find some science-related careers interesting, whether they would consider such careers, and whether they would see themselves in specific occupations.[17] We also collected information on

[16]The structure of the questionnaire could potentially influence students' response rate and answers. To mitigate such biases, we randomly assigned the order of several of the response items (e.g., mathematics/French, man/woman).

[17]We asked students if they would imagine themselves working in specific science-related occupations,

students' perceptions of science-related careers and of gender roles in such careers. The respondents were asked to indicate their level of agreement with five general statements on science using a 4-point Likert scale between 1 (totally agree) to 4 (totally disagree), and to give their opinion regarding five other statements on women in science using a 4-point scale (1 'true'; 2 'kind of true'; 3 'kind of wrong'; and 4 'wrong'). The indexes that we constructed from students' responses to the survey are described in Section 4.

The questionnaire assigned to students in the treatment group included a final section asking them whether they had discussed the classroom intervention with their classmates as well as with their schoolmates from other classes, as a means to indirectly assess the magnitude of potential spillovers. Students in the control group received a slightly different version of the questionnaire for this final section, with no explicit mention of the program.

Students in the treatment group could potentially be more involved ex post, and hence more willing to fill out the questionnaire, typically if the teacher who attended the visit was also the one present in class when the survey was conducted. Table D3 in the Appendix indicates that the survey response rates were high among both Grade 10 students (88 percent) and Grade 12 students (91 percent). The response rates turn out to be slightly higher among Grade 10 students in the treatment group compared to the control group (by 2.6 percentage points). Despite this small difference in response rates, Table D4 in the Appendix shows that the characteristics of Grade-10 students who responded to the survey are generally balanced across the treatment and control groups. The opposite is observed in Grade 12: the survey response rates are similar in the treatment and control groups, but the characteristics of survey respondents exhibit some small but statistically significant differences across the groups, with a slightly larger fraction of high-SES and of non-repeating students in the treatment group compared to the control group. Reassuringly, we show that controlling for students' observable characteristics hardly changes the estimated treatment effects, and hence that our results based on students' responses to the survey are not contaminated by these small imbalances.

some in STEM (e.g., computer scientist, engineer, renewable energy technician, or industrial designer), some in non-STEM (e.g., pharmacist, doctor, chemist, or researcher in biology), and in non-science-related occupations such as therapist or lawyer.

Administrative data. Students' socio-demographic characteristics and enrollment status in academic years 2015-2016 and 2016-2017 are obtained from the *Base Élève Académique* (BEA). These administrative registers, which were provided by the three educational districts of Paris, Créteil and Versailles, cover the universe of students enrolled in the public and private high schools operating in these districts, as well as students enrolled in post-secondary programs located in high schools. They contain basic information on students' demographics (gender, age, country of birth), their social background (measured by parental occupation), and detailed information on their current school enrollment status (school and class attended, elective courses taken). We complement this information with comprehensive individual examination results from the *Diplôme National du Brevet* (DNB), an exam which students take at the end of middle school. We use students' grades on the French and math final exams (converted into percentile ranks), as these exams are externally and anonymously graded at the national level.

For students in Grade 10, enrollment outcomes in the year following the intervention can be measured using the BEA data for 2016-2017. To track the enrollment outcomes of Grade 12 students, we combine two data sources. Students who, after high school graduation, enrolled in selective undergraduate academic programs (CPGE) or vocational training programs (STS) are recorded in the BEA data for 2016-2017, since CPGE and STS programs are physically located in high schools. To track Grade 12 students' enrollment outcomes in other non-selective university undergraduate programs (*Licence*) or in selective vocational programs located in universities (IUT), we use a separate administrative data source (SISE), which is managed by the Ministry in charge of Higher Education. This dataset records all students registered in the French higher education system outside of CPGE and STS, except for the small fraction of students enrolled in paramedic schools. These combined datasets enable us to precisely map where the students in the experimental sample ended up in the academic year following the interventions. We further supplement the enrollment information for Grade 12 students with data from the *Baccalauréat* national exams, which students in the experimental sample took at the end of the academic year 2015-2016.

3.2 Empirical Strategy

Under perfect compliance with random assignment, our experimental design would allow us to directly estimate average treatment effects (ATE) of the program by comparing the outcomes of students in the treatment and control groups. Compliance with the initial random assignment was, however, not perfect since approximately 5 percent of the Grade 10 and Grade 12 classes that were assigned to the treatment group did not receive the visit of a facilitator, while 2 percent of the control classes ended up being visited (see Table C2 in the Appendix).[18]

To deal with this two-way non compliance, we follow the standard practice of using the random assignment as an instrument for actual treatment, which allows us to estimate the local average treatment effect (LATE) instead.[19] Specifically, we estimate the following model using two-stage least squares (2SLS):

$$Y_{ics} = \alpha + \beta D_{ics} + \theta_s + \epsilon_{ics} \tag{1}$$

$$D_{ics} = \gamma + \delta T_{ics} + \lambda_s + \eta_{ics} \tag{2}$$

where Y_{ics} denotes the outcome of student i in classroom c and high school s; D_{ics} is a dummy variable indicating whether the student's classroom actually received the visit of a female facilitator; T_{ics} is a dummy variable indicating assignment to the treatment group. School fixed effects, θ_s and λ_s, are included to account for the fact that the randomization was stratified by school.

The model is estimated separately by grade level and by gender. The error terms ϵ_{ics} and η_{ics} are treated as consisting of a common class effect and an idiosyncratic individual component, and standard errors are clustered at the unit of randomization (classroom level). For individual outcomes within each family of outcomes, we provide adjusted

[18]The direct involvement of two of the authors in the management of the experiment makes us confident that non-compliance was mostly due to organizational and logistical issues rather than an endogenous response to randomization. Female facilitators who ended up carrying out interventions in control classes or in classes that had not been selected to participate in the experiment generally reported that the corresponding classroom interventions had been poorly organized at the school level, with the person in charge often not being aware of the purpose of the visit. In some cases, classroom interventions were scheduled during another specialty course for which classes were mixed, implying that only some of the students in the treatment group were effectively treated. An option could have been to simply drop high-schools were such problems happened. Instead, we adopt a more conservative approach to deal with non-compliance.

[19]Intention to treat (ITT) estimates, which are not reported in the paper, are very close to the LATE estimate due to almost perfect compliance. They are available upon request.

p-values (q-values) in addition to the standard p-values to account for multiple hypothesis testing within each block of outcomes.[20] The method we use is the False Discovery Rate (FDR) control, or the expected proportion of all rejections that are type-I errors.[21] Specifically, we use the sharpened two-stage q-values introduced in Benjamini et al. (2006) and described in Anderson (2008).

Finally, we argue that the Stable Unit Treatment Value Assumption (SUTVA) holds in our setting given that, in the French education system, the class is the unit within which most peer interactions take place. If spillovers do exist, we would only underestimate the true impact of the treatment—an issue which we address in more detail in Section 5.

4 Impact of Classroom Interventions on Student Perceptions, Self-concept and Educational Choices

For female and male students in Grade 10 and Grade 12 (science track) separately, we analyze the impact of the program on three main sets of student outcomes: (i) general perceptions regarding science-related careers and gender roles in science; (ii) preferences and self-concept; and (iii) educational choices. The first set of outcomes captures students' representations of science studies and careers, as well as beliefs regarding the underrepresentation of women in science and its possible causes. These perceptions were directly targeted by the program since the classroom interventions were designed to convey non-stereotyped information regarding science-related careers and women in science. We then investigate if the program affected students' self-concept by changing their preferences regarding science subjects taught at school, their interest in science-related occupations, their self-confidence in math, and how they relate to science in general.[22] Finally, we analyze the program's impact on educational choices.

[20]The families of outcomes we consider are detailed in Section 4 and are labeled as "Negative perceptions of science-related careers," "Taste for science subjects," "Interest in science-related careers," and "Low math self-concept."

[21]The FDR control involves a p-value adjustment less severe than some other methods such as the Familywise Error Rate control or the Bonferroni correction, as long as one is willing to tolerate some type-I error in exchange for a less stringent adjustment.

[22]The results discussed in this section are obtained from empirical specifications that do not control for students' observable characteristics. In Appendix Table G10, we show that the program's estimated effects on student perceptions hardly change when we include such controls, mitigating potential concerns about the small imbalances between the treatment and control groups when considering the sample respondents to the student survey (see Appendix Tables D3 and D4 and the discussion in Section 3.1).

4.1 Effects on General Perceptions of Science-Related Careers and of Gender Roles in Science

Perceptions of science-related careers. The five statements on science careers with which students were asked to agree or disagree concern wages, the length of studies leading to these careers, work-life balance, and the two commonplace stereotypes that jobs in science are dreary and solitary. We build a synthetic index of "negative perceptions of science-related careers" by re-coding the Likert scales such that higher values correspond to a stronger prevalence of stereotypes, before taking the average of each student's responses to the five questions. To facilitate the interpretation of results, we normalize the index to have a mean of zero and a standard deviation of one in the control group. Towards a closer investigation of the various aspects that might be captured by the overall index, we further construct binary variables taking the value one if the student either strongly or slightly agrees with each statement, and zero if she either strongly or slightly disagrees with the statement.[23]

Students' baseline perceptions of science-related careers are indicative of relatively widespread negative stereotypes (see Table 3, columns 1 and 4).[24] More than 80 percent of Grade 10 students in the control group declare that science studies are necessarily long and about 30 percent are not aware of the positive wage premium attached to science-related jobs. A similar fraction of students considers that jobs in science are dreary or solitary and that they are hard to reconcile with a family life. Grade 12 students in the science track have slightly more positive perceptions, but the differences seem limited when balanced with the fact that that these students have already self-selected into the science track in Grade 11.

Overall, the relatively widespread negative or stereotyped perceptions of jobs and careers in science leave room for improvement and may contribute to the large impacts of the program that we find along this dimension. As shown in Table 3, the program significantly improved girls' and boys' perceptions of science-related careers as measured by the synthetic index, in both Grade 10 and Grade 12. Effects range from 15 percent of a standard deviation for boys to around 30 percent of a standard deviation for girls, with

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

[24]These perceptions are described in more detail in a companion paper (Breda et al., 2018).

significantly larger effects for girls. Significant impacts of the program are observed for each individual component of the synthetic index, except for “Wages are higher in science jobs” in Grade 10 and “Jobs in science are dreary” in Grade 12, and are consistent with the program improving students’ general perception of science-related careers. The largest effects are found for the statements “Studies in science are necessarily long” and “Jobs in science are solitary,” which embed two key stereotypes that were explicitly targeted in the slides and videos used by the facilitators. The effects are not strikingly different across gender and grade levels, although they tend to be larger for girls in Grade 12.

Perceptions of gender roles in science. Female underrepresentation in STEM can be broadly attributed to three possible causes: gender differences in abilities, gender discrimination (on the demand side), and gender differences in preferences and career choices (on the supply side). Our survey questions were explicitly designed to capture students’ opinions regarding these three dimensions.

Table 4 reveals the striking fact that over a third of Grade 10 students and a quarter of Grade 12 (science track) students in the control group are not aware that women are underrepresented in science-related careers. These proportions are very similar across gender and grade levels, which is somewhat unexpected as one may have anticipated girls pursuing the science track in Grade 12 to be better informed of female underrepresentation in STEM fields. For both girls and boys in Grades 10 and 12, we find that the program increased awareness of the gender gap in science by 12 to 17 percentage points. This is, perhaps unsurprisingly, one of the outcomes for which the program had the largest effect.

The classroom interventions were also effective at convincing students that women and men have equal aptitude for mathematics. While 19 percent of girls (30 percent of boys) in Grade 10 and 16 percent of girls (26 percent of boys) in Grade 12 believe that men are more gifted than women in math in the control group, these proportions decrease by 2.5 to 5 percentage points in the treatment group.

Regarding the two other broad explanations for the underrepresentation of women in science, the program had more ambiguous—and somewhat unintended—effects. When asked about gender differences in preferences, the share of students who report that women like science less than men is very low in the control group (7 percent of girls and 15 percent of boys in Grade 12), but increases substantially due to the program for both girls and boys, by approximately 30 percent in Grade 10 and 50 percent in Grade 12. The baseline share of girls and boys declaring that women are discriminated in science is much larger (around 60 percent), and it also increases by 7 to 15 percentage points for both gender as a result of the program.

How to interpret these contrasted effects of the classroom interventions on students' perceptions regarding gender roles in science? The key message conveyed by the program materials is that women are strongly underrepresented in science despite the fact that they have the same abilities as men and that science tends to offer better career opportunities.[25] As a consequence, the program slides and videos made salient the underrepresentation of women and discarded gender differences in aptitudes as a cause for the gender gap while remaining agnostic on other possible causes. Since facilitators were left with a lot of freedom and extra time for their interventions, they might have conveyed other messages or shared personal experiences regarding gender discrimination in science careers, which could be an explanation for why their interventions reinforced the belief that women are discriminated in these careers. It seems highly unlikely, however, that female scientists invited to act as role models would have directly conveyed the message that women like science less than men. A more plausible interpretation is that the program's unintended effects on students' perceptions of gender differences in science arose as an attempt to rationalize why there are so few women in STEM careers, making students more likely to agree with the simplistic view that "women like science less than men" and to believe that women are discriminated in these careers.

By making salient the underrepresentation of women in science and by highlighting some of its possible (negative) causes, the program may have limited its potential to steer girls towards science-related careers. This type of phenomenon is likely to extend to other contexts. Role models could have deterring effects towards uninformed subjects by raising their awareness of a potential problem: while they serve as a counter-example to the stereotype, they also make the issue of underrepresentation more visible, which could be counterproductive for some individuals. This point is discussed in more detail in the next section.

[25] The second video shown in class ("Are we all Equal in Science?") was designed to convince students that women and men have similar brains and are equally capable of succeeding in math and science.

4.2 Effects on Stated Preferences and Self-Concept

We now turn to the effects of the program of students' stated preferences and self-perception. Specifically, we investigate how the classroom interventions affected boys and girls' taste for science subjects, their interest in science-related careers, and their self-concept in math. Table 5 reports the estimated treatment effects along these three dimensions, based on the summary indices that we constructed from students' answers to the questionnaire. Treatment effects for each of the questions being aggregated in these indices are reported separately in Appendix Tables G7, G8, and G9.

Taste for science subjects. The program had no significant effect on students' taste (reported on a 0 to 10 Likert scale) for the scientific subjects taught at school, i.e., math, physics-chemistry, and biology-geoscience, nor on their self-reported taste for science in general (see Appendix Table G7). This lack of effects is also found when using the standardized index that aggregates students' responses to the four questionnaire items (see Table 5). These findings are not surprising, as the program did not expose students to science-related content and was not specifically designed to promote interest in science subjects.

Interest in science-related careers. Choosing a science career does not only depend on students' taste for the science subjects taught at school. It also depends on their perceptions of science-related jobs and the type of amenities they may provide, such as wages, work/life balance, working environment, etc. As the program is found to have significantly improved students' perceptions of science careers, one might expect students in the treatment group to be more likely to consider these careers for themselves, even if their taste for science subjects was unaffected. We find evidence of such effects for girls in Grade 12, since their interest in science-related careers appears to have increased by approximately 10 percentage of a standard deviation due to the program. Interestingly, the more detailed results reported in Appendix Table G8 show that girls in Grade 12 are the only subgroup of students to have become more aware that science jobs pay well. They are also more likely to declare that expected future earnings are important for their choice of study. However, when faced with a list of seven science-related and three non-science-related jobs—including some jobs that were explicitly mentioned in the materials used by facilitators—none of the groups of students appears significantly more likely to declare a stronger preference for either of these jobs.[26]

Math self-concept. Consistent with the literature, our sample exhibits large gender differences in math self-concept, both in Grade 10 and in Grade 12 (see Appendix Table G9). In the control group, girls' self-reported math performance in Grade 10 is 15 percent of a standard deviation lower than that reported by boys, and 10 percent of a standard deviation lower in Grade 12. Moreover, girls are 15 to 20 percentage points more likely to declare being lost in front of a math problem, and to be worried when thinking about math.

Despite being a light-touch intervention, the program had moderate positive effects on students' math self-concept. Using the index constructed from students' answers to the three math self-concept questions, positive effects are only found to be statistically significant for boys in Grade 12. The program, however, appears to have reduced the probability of reporting being worried when thinking about math for all groups of students.[27] Point estimates tend to be larger for boys than for girls both in Grade 10 and Grade 12, implying that the program had not effect on the large gender gap in math self-concept and rather left it unchanged. These results are in contrast with our previous findings regarding the effects on students' perceptions of science-related and their interest in such careers, which we found to be stronger for girls than for boys.

4.3 Effects on Educational Choices

Choices of high school track after Grade 10. We find no significant effects of the program on the educational choices of Grade 10 students in the academic year following the intervention, i.e., 2016-2017. Analysis of the administrative enrollment data indicates that both girls and boys in the treatment group are not more likely to be enrolled in a general or technical science track in Grade 11 than students in the control group (see Table 6, Panel A). Estimated treatment effects are all very close to zero, irrespective

[26] Point estimates are of the expected sign, but effects tend to be too small to be detectable except in Grade 10, where female students in the treatment group are 2 to 3 percentage points more likely to report considering a career as an engineer or an industrial designer (results available upon request).

[27] For each group of students, correction of p-values for multiple testing cannot rule out that the effects are driven by chance alone, with the caveat that finding a significant effect for the same variable in each of the four groups seems an unlikely event.

of whether we consider the probability of attending any STEM track or whether we distinguish between general and technical STEM tracks. These results imply that the program did not significantly reduce the 20 percentage points gender gap in the likelihood of pursuing STEM studies after Grade 10. We find similar results when considering the study intentions reported by Grade 10 students in the post-treatment survey, suggesting that the lack of effects on enrollment outcomes is due to students' choices being unaffected rather than to schools being less likely to admit treatment group students in the science track. Altogether, these results are consistent with the previous finding that Grade 10 students' interest in science-related careers was essentially unaffected by the program.

Choice of college major after Grade 12. One of the central results of this paper is that the program had significant effects on the educational choices of female students in Grade 12, while having no sizable effects on the choices of their male counterparts. The enrollment outcomes of Grade 12 students can be measured with a high degree of accuracy from the various administrative registers that cover almost all students in higher education (see Section 3). Panel B of Table 6 shows that among girls in the treatment group, the probability of being enrolled in a STEM program after graduating from high school increased by 2.4 percentage points (significant at the 10 percent level), which represents a 10 percent increase from a baseline of 28 percent. The corresponding increase for boys is close to zero. As a result of this differential response to the intervention, the gender gap in STEM enrollment in higher education was reduced to 16.5 percentage points from a baseline of 18.6 percentage points, i.e., a 10 percent reduction.

This aggregate impact is driven by a significantly larger number of girls enrolling in (i) selective STEM programs and (ii) the most male-dominated STEM programs, i.e., math, physics, and computer science. These results are particularly striking since selective and male-dominated STEM programs are not only the most prestigious tracks but also those with the largest initial gender gap in enrollment.

Boys who pursue STEM studies after high school graduation are equally likely to enroll in selective and non-selective programs, whereas girls are 70 percent more likely to enroll in non-selective STEM programs. As a consequence, the gender gap in selective STEM programs is much larger than in the non-selective ones. It is therefore remarkable that the classroom interventions led to a significant 3.5 percentage point increase in the fraction of female students enrolling in selective STEM programs, which represents a 30 percent increase from a baseline enrollment rate of 11 percent. The corresponding estimates for male students suggest that the classroom visits may also have increased boys' enrollment in these programs, but the effects are not statistically significant.[28] Comparing the point estimates for girls and boys to the respective baseline rates would imply a 10 to 20 percent reduction in the gender enrollment gap in selective STEM due to the classroom visits.

By construction, there is almost no gender gap in non-male-dominated STEM undergraduate programs. These programs concern 12 percent of girls and 9 percent of boys in our sample. By contrast, the fraction of students enrolling in male-dominated STEM programs after Grade 12 is 16 percent among girls and 38 percent among boys in the control group. This gap was reduced by the classroom interventions since the proportion of girls enrolling in such programs increased by a statistically significant 3.8 percentage points (i.e., a 20 percent increase from the baseline), compared to a non-significant 1.7 percentage point increase for boys.

Our estimates imply that, on average, the classroom visits induced one girl in every two science-track Grade 12 classes to switch to a selective or a male-dominated STEM program when entering higher education.[29] The magnitude of the effects can also be assessed by evaluating how the intervention would affect the share of female students in the different types of STEM programs if our estimates were extrapolated without considering general equilibrium effects. A simple back-of-the-envelope computation shows that this share would increase from 35 to 37 percent in STEM programs altogether, from 30 to 34 percent in selective STEM programs, and from 28 to 31 percent in male-dominated STEM programs.

Considered together, the results for girls in Grade 10 and Grade 12 indicate that the program was only effective at steering girls towards the STEM tracks where there are initially strongly underrepresented. This is despite the fact that two thirds of the role models come from fields where women are not strongly underrepresented (earth and life science) and that the program was designed to promote all types of STEM careers,

[28] Considering only the most selective STEM programs (CPGE) and leaving aside the less selective ones (STS and IUT) yields a slightly larger and statistically significant effect for boys (after correcting for multiple hypothesis testing), suggesting that the program may have shifted boys across STEM programs, from the least to the most difficult and selective.

[29] This computation is based on an average of 15 girls per class and an estimated 3.5 (respectively 3.8) percentage point increase in the probability of enrolling in a selective (respectively male-oriented) STEM program.

including those where women now outnumber men—at least in France. These results suggest that in the current setting, role models are only affecting the most stereotyped choices. Promoting a choice that is no longer associated with strong gender norms—such as studying biology or medicine nowadays for female students—has no effect, at least not through a one-hour intervention.[30]

5 Persistence of Effects, Timing of Visits, and Spillovers

This section extends the analysis to the persistence of effects, the role of the timing of interventions, and potential spillovers.

Persistence. The effects of the program on students' perceptions that we measure between two and six months after the classroom visits may well be short-lived. To check this, Appendix Table I11 investigates how treatment effects vary depending on the time interval between the intervention and the post-treatment survey, which we spit into three groups: one to two months, three to four months and five to six months.

The limited sample size for each time interval—especially when considering students who were surveyed five to six months after the visit—and the possibility that the quality of facilitators' interventions may have changed over time are two potential limitations that call for caution in drawing firm conclusions about the persistence of effects. With these caveats in mind, the results suggest that the treatment effects do not vanish quickly over time, as most of them remain statistically significant four months after the interventions. In all cases, the program's effects lasted long enough to affect college choices in Grade 12.

Timing and order of visits. If anything, earlier interventions had larger effects on Grade 12 students' higher education choices, which are made in April (see Appendix

[30]The absence of effects on female students' educational choices in Grade 10 may be explained by the fact that opting for the science track in Grade 11 is the most common choice. It is not strongly associated with having a strong inclination towards science in part because more than half of the students in this track do not pursue in a science major after graduating from high school. In fact, many undecided yet high-performing students tend to choose the science track in high school because it will give them access to most fields of study in higher education. As shown in Section 6, the program's effects on girls' educational choices in Grade 12 are driven by the best students, and there is probably not much room to affect these students' choices in Grade 10 because those who turn away from the science track are already making a strong choice, which is difficult to reverse.

Figure J8). For girls, classroom visits that took place in November increased by 10 percentage points the likelihood of enrolling in a selective STEM program after high school graduation, whereas the effects of visits in December or January are estimated to be closer to 5 percentage points, and not statistically significant for visits in February and March.[31] These findings provide suggestive evidence that interventions scheduled when many students are still undecided about their major and future career path can be more effective than those taking place just before irreversible choices are made, at a time when most students have settled on a decision.

Facilitators carried out an average of five to six classroom interventions, which were typically scheduled during two half-day visits in two distinct schools. Whether the quality of their interventions could have changed over time is a priori unclear. On the one hand, it is possible that repeated interactions with students could have made facilitators more effective at influencing students' choices; on the other hand, visiting several classroom in a row might have been negatively affected their performance because of tiredness. We find no evidence that the effects on enrollment outcomes are systematically related to the order in which the classrooms were visited in a given school (see Panel A of Appendix Figure J9). Facilitators, however, seem to have been more effective at affecting students' educational choices during their first visit to a school, typically consisting of three consecutive classroom interventions, than during the last visit to a school (see Figure J9, Panel B). For boys, this pattern is statistically significant and not driven by the fact that facilitators' first visit to a school took place earlier in the academic year than their third visit, and that earlier interventions tend to be more effective.[32] Lastly, ordering all interventions by their date from 1 to 7, we fail to detect any clear pattern over time (Figure J9, Panel C).

Overall, these results suggest that facilitators' performance did not systematically improve or worsen over time, possibly because their impact on student outcomes has more to do with the embodiment of a science career (a pure role model effect) than with their effectiveness at conveying information, which is more likely to have improved

[31] The difference in the estimated effects of visits that took place before vs. after February 1 is statistically significant at the 5 percent level for girls and at the 10 percent level for boys. These differences are robust to controlling for a possible improvement or decline in the quality of facilitators' interventions over time (see next paragraph).

[32] The difference between the effects of enrollment outcomes of the first vs. the third visit to a school remains statistically significant when controlling for the the month in which the visit took place.

over time. Our results on the trends in facilitators' performance should nevertheless be considered with caution, for the same reasons as the results on the persistence of effects. Indeed, high schools that were the first to be visited are also those that agreed early to participate in the experiment and/or those that were chosen first by facilitators. These schools might therefore have had a greater potential for large treatment effects.

Spillovers. An important issue to consider is the extent to which the program could have induced spillover effects for students in the control group. These students may have heard about the visits either directly, through their schoolmates in treatment group classes, or indirectly, through regular social interactions. Information about the classroom interventions might also have been disseminated to students in te control group by teachers who were present on the day of the facilitator's visit to the school—teachers being typically in charge of multiple classes in the same school. Provided that the direction of effects is the same for students in the treatment and control groups, spillovers would be expected to lead to an underestimate of the program's impacts.

To get some sense of the likely magnitude of these spillover effects, the last section of the survey questionnaire asked treatment group students whether they talked about the program with their classmates, with schoolmates from other classes, and with friends from other schools; we also asked students in the control group whether they had heard about a science-related awareness-raising program and, more specifically, whether they knew about other classrooms in the schools receiving the visit of a female scientist.

The student survey data indicate that in the treatment group, 58 percent of Grade 10 students and 63 percent of Grade 12 students report having talked about the program with their classmates, but only 24 percent (27 percent) with schoolmates from other classes, and 20 percent with students from other schools (see details in Appendix Table K12). Interestingly, these proportions are higher for girls than for boys in the treatment group: 66–70 percent of girls in report having discussed the program with their classmates and 28–33 percent with schoolmates from other classes vs. respectively 50–56 percent and 20–21 percent among boys.

In the control group, 86 percent of Grade 10 students report not having heard of a classroom visits in other classes, while 12 percent have vaguely heard of a visit, and 3 percent are well aware of such visits. The gender difference in these proportions is small.

Grade 12 students in the control are more likely to report having heard about visits in other classes (34 percent versus 14 percent, mostly in a vague manner), which may not be surprising as the share of Grade 12 (science track) classes assigned to the treatment group among all Grade 12 (science track) classes in participating schools is typically larger than the corresponding fraction for Grade 10 classes, on average 31 percent vs. 24 percent.

Altogether, the survey evidence points to limited spillover effects between treatment and control classes. To test this more formally, we implement two empirical strategies. A first rudimentary approach consists of comparing the outcomes of students in the control group who report not having heard about the visits to the outcomes of those who report being at least vaguely aware that other classes were visited by a female scientist, controlling for students' social background and academic performance (measured by their percentile rank on the DNB exams in math and French). We find that most of the outcomes that were impacted by the program for students in the treatment group are not significantly different across both types of control group students (see first column of Tables K13 and K14 in Appendix). There are, however, a few significant differences, such as, among girls in Grade 10, their math self-concept or their interest in science-related careers.

An obvious limitation of the above comparison is that students with a stronger inclination for science might also have been more likely to hear and ask about the program, which would confound the comparison. As an attempt to overcome this problem, we use the within-school, within-grade-level share of classes assigned to treatment an instrument for the probability of hearing about the program for students in the control group. What motivates this approach is that the larger the share of classes assigned to treatment among all classes in the same grade level, the higher the chances that a student in the control group would be connected with a schoolmate whose classroom was visited. The first-stage estimates for girls indicate a strong positive correlation between the share of classes assigned to the treatment group and the probability that students in the control group have heard about the program, in both Grade 10 and Grade 12. The 2SLS estimates for girls often change sign and are statistically insignificant for all outcomes in both grades, suggesting that having heard about the program was not sufficient to change the perceptions or behavior of girls in the control group. An important limitation of this approach, howevern is that the 2SLS estimates are very imprecise and hence do not provide definitive evidence against spillover effects. Another limitation is that the instrument is by construction negatively correlated with school size, since most of the between-school variation in the share of classes assigned to treatment in a given grade level is driven by the between-school variation in the total number of classes belonging to that grade level. If students in large vs. small schools differ along dimensions that are not captured by our controls, the exclusion restriction could be violated, leading to biased estimates. To address this issue, we show in the third column of Appendix Tables K13 and K14 that the results are robust to controlling for school size, i.e., the number of classes for the considered grade level in the school. Note, however, that controlling for school size further reduces the precision of estimates. The 2SLS estimates for boys are essentially inconclusive, as the share of classes assigned to the treatment group in a given school and grade level turns out to be only weakly correlated with male students' probability of having heard about the program (columns 4 to 6).

To investigate the possibility that students in the control group could have been indirectly impacted by the program without having necessarily heard about it, we adopt a similar approach. We use the share of classes assigned to treatment in each school and grade level as a proxy for the magnitude of potential spillovers. We then estimate a simple OLS regression of individual outcomes on this share for all students in the control group, irrespective of whether or not they have heard about the program. Controlling for students' socio-demographic characteristics and prior academic performance, we find no systematic evidence of student outcomes being correlated with our proxy for potential spillovers—at least not with the expected sign (see Appendix Tables K15 and K16).

Overall, the survey evidence and the above regression results lead us to conclude that spillovers between treatment and control classes were limited, if any.

6 Mechanisms: How Do Role Models Affect Student Behavior?

This section attempts to shed some light on the reasons why light-touch classroom interventions by female role models with a science background are able to modify girls' choice of study at college entry. Our insights stem from the comparison of groups of students who were exposed to different role models or responded differently to a given role model.

We proceed in two steps, first by investigating how the treatment effects on enrollment vary across different groups of students and, second, by looking at how these students responded to the program along other dimensions. The current analysis is restricted to the two most obvious dimensions of heterogeneity: student academic performance and facilitator background (i.e., professionals privately employed by the firm whose foundation is supporting the program vs. young researchers receiving a fellowship from the foundation). A more systematic analysis of treatment effects heterogeneity using machine learning techniques is still work in progress.

Heterogenous effects by student academic performance. Our empirical analysis has revealed that the program's main effect on educational choices was to induce a significant share of girls in Grade 12 to switch to selective and male-dominated undergraduate STEM programs after graduating from high school. This finding naturally raises the question of how the effects differ across students with different levels of academic performance.

Science majors are in France considered the most competitive and prestigious ones. In high school, admission to the science track in Grade 11 requires achieving sufficiently high grades in science subjects in Grade 10. At college entry, only students with top grades in mathematics, physics and biology-geoscience, stand a chance of being admitted to a selective STEM program. One would therefore expect the enrollment effects observed among girls in Grade 12 to be driven by the best-performing ones. This is indeed what we find when comparing enrollment effects across students of different achievement levels.

We use students' percentile rank on the *Baccalauréat* final exam in mathematics to proxy their academic preparedness for selective STEM programs, since math performance is the single most important criterion that these programs consider when screening applications. As Grade 12 students in the treatment group took the *Baccalauréat* exams a few months after the visits, a legitimate concern is that this measure of academic performance could have been affected by the classroom interventions and might therefore induce endogenous selection bias if used as a conditioning variable. In our setting, this concern does not appear to be justified since, as show in Appendix Table L17, the program had no significant effect on students' *Baccalauréat* final exam grades in math. Although female facilitators could in principal have increased students' motivation to be admitted to the most selective STEM programs, we find no evidence that the intervention incentivized students to increase their effort in science subjects, by dedicating more time to mathematics for instance. This lack of effects on student performance is consistent with the previous finding that the program had no impact on students' taste for science subjects taught at school.

We investigate heterogeneity in the program's effects on enrollment outcomes by comparing Grade 12 students in different quintiles of the *Baccalauréat* performance in mathematics. The results, which are displayed in Figure 2, show that the program's positive impact on selective STEM enrollment for girls in Grade 12 is mostly driven by girls in the top quintile of math performance. For these female students, the probability of being enrolled in a selective STEM program one year after high school graduation increased by 50 percent from a baseline of 28 percent.[33] While the program also appears to have induced some male students in the top quintile of math performance to enroll in selective STEM programs, the effect is smaller and only marginally significant (18 percent increase from a baseline of 50 percent). These findings are consistent with top-performing girls in math being probably the closest to the margin between applying and not applying to selective STEM undergraduate programs, and hence the most likely to change their mind in response to the intervention. Especially striking is the fact that among the 20 percent top achievers in math, the treatment reduces by a third the large baseline gap between girls and boys' enrollment in selective STEM programs, from 22 percentage points to 14 percentage points.

To shed light on the potential mediators explaining why the classroom interventions were more effective at steering high-achieving girls towards selective STEM programs, we compare in Table 7 the effects of the program on boys and girls' perceptions by splitting the Grade 10 and Grade 12 samples into two groups according to student performance in math. While the results do not point to large and significant differences between high- and low-achieving students, they do suggest that students with above-median performance in math were more responsive to the program's "positive" messages (e.g., studies in science are not necessarily long, women and men have equal aptitude for math) and were more likely to show increased interest in science-related jobs than students with below-median

[33]We have chosen to focus on enrollment in selective STEM for the heterogeneity analysis as this is the enrollment outcome for which we find the largest LATE. All results are however qualitatively similar when we look instead at enrollment in male-dominated STEM or in all types of STEM tracks. Detailed results are provided in Appendix Table M18 and Figures H6 and H7.

performance. By contrast, the classroom interventions seem to have induced a larger share of low-achieving student to declare that women are discriminated in science careers. These findings provide suggestive evidence that the program's unintended effect of raising students' awareness of female underrepresentation in science could have mitigated its ability to convince low-achieving girls of the benefits of such careers for themselves.[34]

Heterogenous effects by facilitator background. Previous research suggests that the effect of role models could vary according to how group members perceive their own ability, and to how personally relevant the role model is to them (Lockwood and Kunda, 1997). In this respect, the characteristics of role models could matter (Nguyen, 2008). In our setting, we provide suggestive evidence that the professional background of female facilitators was differently relevant to different students.

Although the assignment of facilitators to participating high schools was not subject to random assignment, Table F6 in the Appendix indicates that the observable characteristics of classrooms assigned to the treatment group are reasonably balanced according to whether the facilitator is an employee of the firm whose foundation is supporting the program or a researcher (Ph.D. or postdoc) receiving a fellowship from the foundation. It seems therefore reasonable to assume that significant differences in the estimated effects of both types of facilitators on student outcomes are unlikely to be confounded by unobservable differences in the classrooms they visited.

As shown in Table 2, facilitators with a research background are on average younger, which may foster a stronger sense of identification by the students. However, because they work in highly specialized fields and in very competitive environments, it is not clear how attainable students might consider the achievements of such role models. On the other hand, professionals working for the sponsoring firm have on average higher wages, more experience, and come from a less purely academic background. They also hold a permanent position contrary to Ph.D.s and postdocs. How these different types of role models may affect student attitudes and behavior is therefore a priori ambiguous.

Figure 3 shows that the female professionals had a large and significant 5.2 percentage points effect on the probability of enrolling in a selective STEM program for girls in

[34] The subgroup analysis in Table 7 confirms that girls with above-median performance in math are significantly more likely to enroll in a selective STEM program as a result of the program than their below-median counterparts ($p = 0.015$)

Grade 12, whereas researchers had no detectable effect on the educational choices of both female and male students—the difference between the effects of both types of facilitators being statistically significant at the 5 percent level (the corresponding point estimates and p-values are reported in Appendix Table M19). Estimated effects for boys are also stronger among those whose classroom was visited by a professional, although the point estimates are not statistically significant at conventional levels.

Comparing the effects of professionals and researchers along other outcomes, we find that both types of facilitators were equally competent at changing boys and girls' general perceptions regarding science-related careers and gender roles in science. The results reported in Table 8 suggest, however, that female professionals were more effective at convincing girls in Grade 12 to consider STEM careers for themselves. The larger effect of professionals on the educational choices of this group of students could have have mediated by the fact that they made girls more aware of the wage premium associated with science-related careers, and of the importance of considering this dimensions when choosing a career path. These results suggest that the ability of role models to affect girls' educational choices requires not only challenging gender-stereotyped views about STEM careers, but also personifying an attractive and attainable pathway to such careers.

7 Conclusion and Discussion

Based on a large-scale randomized field experiment, this paper provides empirical support to the hypothesis that light-touch in-class intervention by external female role models are effective at mitigating gender stereotypes among high school students and—at least for the high achievers in math—have the ability to influence girls' educational choices.

We first document gender differences in attitudes toward science-related careers, as well as the prevalence of stereotypical opinions regarding gender roles in science among high school students. Using a random assignment of a class-based intervention to students in Grade 10 and in Grade 12 (science track)—which in France are the two critical turning points in students' educational careers—to a one-hour intervention, we investigate the causal impact of role models on aspirations, attitudes, and educational choices. We find that external female role models significantly reduce the prevalence of stereotypes associated to jobs in science, for both female and male students, as well as stereotypes related to innate gender differences in cognitive abilities. However, the intervention simultaneously increases the salience of the underrepresentation of women, and therefore the beliefs that women have a less pronounced taste for science and that they are discriminated in STEM careers. These findings suggest that students rationalize gender segregation among occupations as reflecting differences in (potentially socially constructed) tastes or discrimination. Despite these somewhat unintended effects, the role models did increase students' propensity to projects themselves into science-related careers.

Using administrative data for the follow-up year, we provide evidence that the program had a significant effect on the educational choices of girls in Grade 12 (science track), without affecting the enrollment outcomes of students in Grade 10 and only marginally of boys in Grade 12. Our estimate indicates that the program increased by 10 percent the fraction of female students in Grade 12 enrolling in STEM undergraduate program. These effects are driven by high-achieving girls in math switching to selective and male-dominated STEM programs, which is consistent with these students being the closest to the margin between STEM and non-STEM pathways in higher education.

We provide suggestive evidence that role models' characteristics matter for effectiveness. In particular, we find that female facilitators working in the private sector were more effective than young researchers at steering girls in Grade 12 towards STEM undergraduate programs after high school graduation, despite the fact that both types of role models were equally competent at improving students' general perceptions of STEM careers and at reducing gender stereotypes. The finding that professionals had generally stronger effects on students' views about science-related careers than researchers, particularly among female students in Grade 12, suggests that their career pathways may have been perceived as more attainable and relevant to the students, which could explain why they had larger impact on the decision to pursue STEM studies.

Our results indicate that the female role models participating in the program were only able to significantly reduce the gender gap in undergraduate STEM enrollment among students with a strong academic background and who selected the science track in high school. More research is needed to investigate whether role models with more diverse backgrounds (e.g., in terms of their age, ethnicity, educational attainment, or occupation) could be more relevant to lower-achieving female students and could influence their educational choices earlier in life.

References

Anderson, Michael L., "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*, 2008, *103* (484).

Babcock, Linda and Sara Laschever, *Women don't Ask: Negotiation and the Gender Divide*, Princeton University Presss, 2012.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova, "Female Leadership Raises Aspirations and Educational Attainment for Girls: A policy Experiment in India," *Science*, 2012, *335* (6068), pp. 582–586.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli, "Adaptive Linear Step-up Procedures that Control the False Discovery Rate," *Biometrika*, 2006, *93* (3), 491–507.

Bettinger, Eric P. and Bridget Terry Long, "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students," *American Economic Review*, 2005, *95* (2), pp. 152–157.

Betz, Diana E. and Denise Sekaquaptewa, "My Fair Physicist? Feminine Math and Science Role Models Demotivate Young Girls," *Social Psychological and Personality Science*, 2012, *3* (6), pp. 738–746.

Black, Dan A., Amelia M. Haviland, Seth G. Sanders, and Lowell J. Taylor, "Gender Wage Disparities among the Highly Educated," *Journal of Human Resources*, 2008, *43* (3), 630–650.

Blau, Francine D. and Lawrence M. Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Breda, Thomas and Mélina Hillion, "Teaching Accreditation Exams Reveal Grading Biases Favor Women in Male-Dominated Disciplines in France," *Science*, 2016, *353* (6298), pp. 474–478.

_ **and Son-Thierry Ly**, "Professors in Core Science Fields are not always Biased Against Women: Evidence from France," *American Economic Journal: Applied Economics*, 2015, *7* (4), pp. 53–75.

_ **, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre**, "Les filles et les garcons face aux sciences : Les enseignements d'une enquête dans les lycées franciliens," *Éducation et Formation*, 2018, *97*, 1–35.

Brown, Charles and Mary Corcoran, "Sex-Based Differences in School Content and the Male-Female Wage Gap," *Journal of Labor Economics*, 1997, *15* (3), pp. 431–465.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek, "Gender, Competitiveness, and Career Choices," *Quarterly Journal of Economics*, 2014, *129* (3), 1409–1447.

Canes, Brandice J. and Harvey S. Rosen, "Following in her Footsteps? Faculty Gender Composition and Women's Choices of College Majors," *Industrial and Labor Relations Review*, 1995, *48* (3), pp. 486–504.

Card, David and Abigail Payne, “High School choices and the Gender Gap in STEM,” 2017. NBER Working Paper No. 23769.

Carlana, Michela, “Implicit Stereotypes: Evidence from Teachers’ Gender Bias,” 2018. IZA Working Paper No. 11659.

Carnevale, Anthony P., Nicole Smith, and Michelle Melton, *STEM: Science, Technology, Engineering, Mathematics*, Washington, DC: Georgetown University Center on Education and the Workforce, 2011.

Carrell, Scott E., Marianne E. Page, and James E. West, “Sex and Science: How Professor Gender Perpetuates the Gender Gap,” *Quarterly Journal of Economics*, 2010, *125* (3), pp. 1101–1144.

Ceci, Stephen J. and Wendy M. Williams, “Understanding Current Causes of Women’s Underrepresentation in Science,” *Proceedings of the National Academy of Sciences*, 2011, *108* (8), pp. 3157–3162.

_ , Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams, “Women in Academic Science: A Changing Landscape,” *Psychological Science in the Public Interest*, 2014, *15* (3), pp. 75–141.

Cheryan, Sapna, John Oliver Siy, Marissa Vichayapai, Benjamin J. Drury, and Saenam Kim, “Do Female and Male Role Models who Embody STEM Stereotypes Hinder women’s Anticipated Success in STEM?,” *Social Psychological and Personality Science*, 2011, *2* (6), pp. 656–664.

Correll, Shelley J., “Gender and the Career Choice Process: The Role of Biased Self-Assessments,” *American Journal of Sociology*, 2001, *106* (6), pp. 1691–1730.

Dasgupta, Nilanjana and Shaki Asgari, “Seeing is Believing: Exposure to Counterstereotypic Women Leaders and its Effect on the Malleability of Automatic Gender Stereotyping,” *Journal of Experimental Social Psychology*, 2004, *40* (5), pp. 642–658.

Dee, Thomas S, “Teachers and the Gender Gaps in Student Achievement,” *Journal of Human Resources*, 2007, *42* (3), pp. 528–554.

Eble, Alex and Feng Hu, “Stereotypes, Role Models, and the Formation of Beliefs,” 2017. CDEP-CGEG Working Paper No. 43.

Ehrlinger, Joyce and David Dunning, “How Chronic Self-Views Influence (and Potentially Mislead) Estimates of Performance,” *Journal of Personality and Social Psychology*, 2003, *84* (1), pp. 5–17.

Gale, David E. and Lloyd S. Shapley, “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, 1962, *69* (1), 9–15.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini, “Performance in Competitive Environments: Gender Differences,” *Quarterly Journal of Economics*, 2003, *118* (3), pp. 1049–1074.

Hoffmann, Florian and Philip Oreopoulos, "A Professor Like Me: The Influence of Instructor Gender on College Achievement," *Journal of Human Resources*, 2009, *44* (2), pp. 479–494.

Hyde, Janet S., "The Gender Similarities Hypothesis," *American Psychologist*, 2005, *60* (6), pp. 581–591.

Lavy, Victor and Edith Sand, "On the Origins of Gender Gaps in Human Capital: Short- and Long-Term Consequences of Teachers' Biases," *Journal of Public Economics*, 2018, *167*, 263–269.

Leibbrandt, Andreas and John A. List, "Do Women Avoid Salary Negotiations? Evidence from a Large-Scale Natural Field Experiment," *Management Science*, 2015, *61* (9), 2016–2024.

Lim, Jaegeum and Jonathan Meer, "The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea," *Journal of Human Resources*, 2017, pp. 1215–7585.

Lockwood, Penelope and Ziva Kunda, "Superstars and Me: Predicting the Impact of Role Models on the Self.," *Journal of Personality and Social Psychology*, 1997, *73* (1), p. 91.

MENESR (Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche), *Filles et garçons sur le chemin de l'égalité, de l'école à l'enseignement supérieur*, Paris: MENESR, 2018.

Muralidharan, Karthik and Ketki Sheth, "Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers," *Journal of Human Resources*, 2016, *51* (2), 269–297.

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

Neumark, David and Rosella Gardecki, "Women Helping Women? Role Model and Mentoring Effects on Female Ph.D. Students in Economics," *Journal of Human Resources*, 1998, *33* (1), pp. 220–246.

Nguyen, Trang, "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar," 2008. Manuscript.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy away from Competition? Do Men Compete too Much?," *Quarterly Journal of Economics*, 2007, *122* (3), pp. 1067–1101.

_ **and** _ , "Explaining the Gender Gap in Math Test Scores: The Role of Competition," *Journal of Economic Perspectives*, 2010, pp. 129–144.

O'Brien, Laurie T., Aline Hitti, Emily Shaffer, Amanda R. Van Camp, Donata Henry, and Patricia N. Gilbert, "Improving Girls' Sense of Fit in Science - Increasing the Impact of Role Models," *Social Psychological and Personality Science*, 2016, *8* (3), pp. 301–309.

OECD, *Education at a Glance 2016*, Paris: OECD Publishing, 2016.

Olson, Steve and Donna Gerardi Riordan, *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, Executive Office of the President: President's Council of Advisors on Science and Technology, 2012.

Paredes, Valentina, "A Teacher Like Me or a Student Like Me? Role Model Versus Teacher Bias Effect," *Economics of Education Review*, 2014, *39*, 38–49.

Porter, Catherine and Danila Serra, "Gender Differences in the Choice of Major: The Importance of Female Role Models," 2017. Manuscript.

Rothstein, Donna S., "Do Female Faculty Influence Female Students' Educational and Labor Market Attainments?," *Industrial and Labor Relations Review*, 1995, *48* (3), pp. 515–530.

Spelke, Elizabeth S., "Sex Differences in Intrinsic Aptitude for Mathematics and Science: a Critical Review," *American Psychologist*, 2005, *60* (9), pp. 950–958.

Terrier, Camille, "Boys Lag Behind: How Teachers' Gender Biases Affect Student Achievement," 2016. SEII Discussion Paper.

Weinberger, Catherine J, "Mathematical College Majors and the Gender Gap in Wages," *Industrial Relations*, 1999, *38* (3), pp. 407–413.

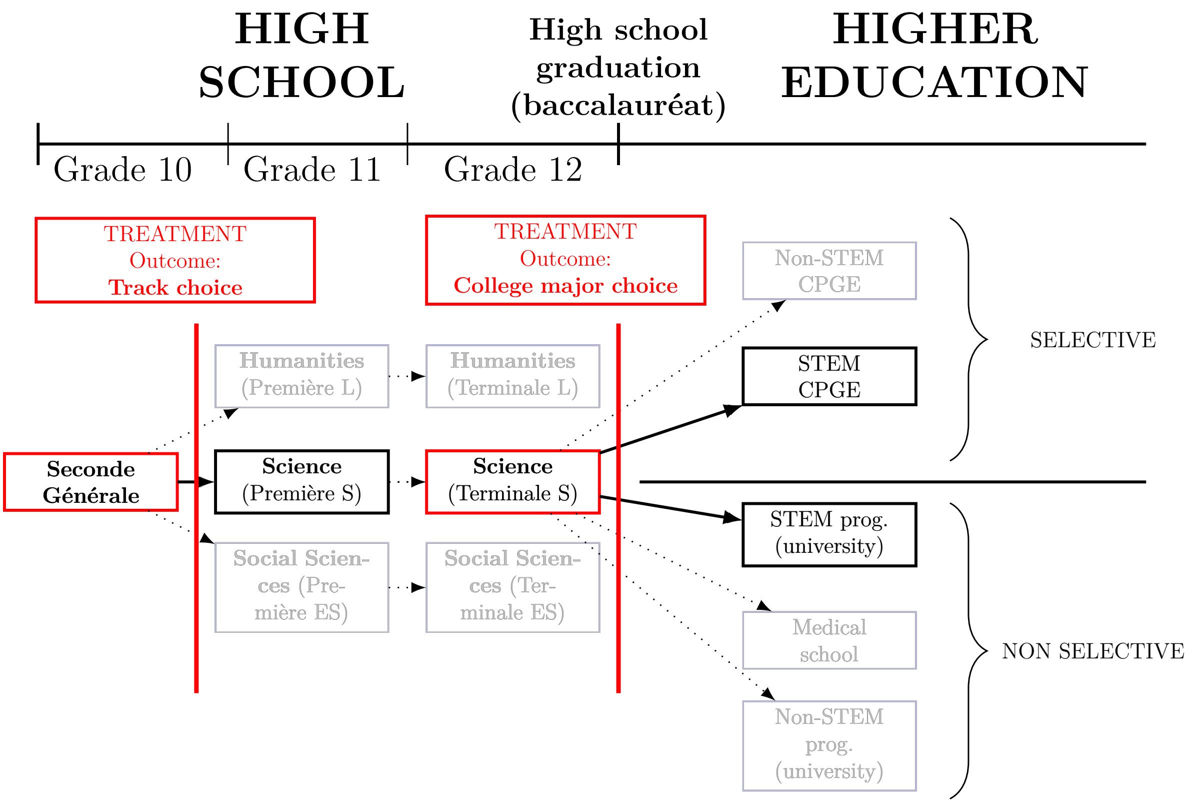
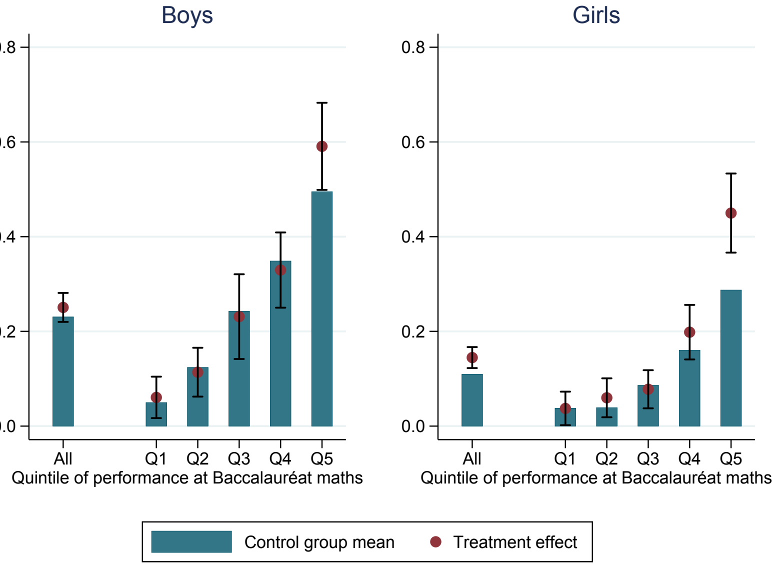
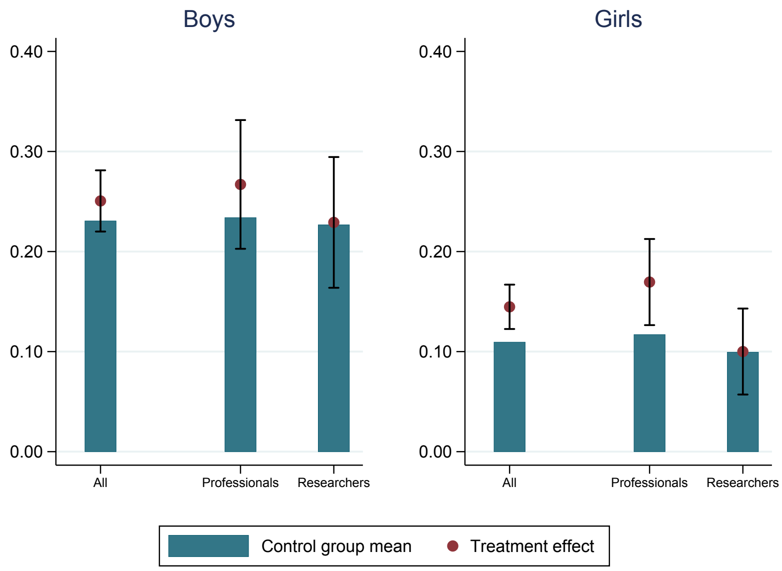


Figure 1: Educational Tracks in Upper Secondary Education in France



Notes: The figure shows the fraction of Grade 12 (science track) students who enrolled in selective STEM undergraduate programs after graduating for high school, for boys (left panel) and girls (right panel) separately. The blue bars indicate the overall mean among students in the control group and the separate means by quintile of *Baccalauréat* performance in math. The red solid dots show the estimated treatment effects with 95 percent confidence intervals denoted by vertical capped bars.

Figure 2: Grade 12 (Science Track) – Enrollment in Selective STEM the Following Year, by Student Gender and Quintile of *Baccalauréat* Performance in Math



Notes: The figure shows the fraction of Grade 12 (science track) students who enrolled in selective STEM undergraduate programs after graduating for high school, for boys (left panel) and girls (right panel) separately. The blue bars indicate the overall mean among students in the control group and the separate means by background of the female facilitator who visited the classroom (professional or researcher). The red solid dots show the estimated treatment effects with 95 percent confidence intervals denoted by vertical capped bars.

Figure 3: Grade 12 (Science Track) – Enrollment in Selective STEM the Following Year, by Student Gender and Background of Female Facilitator

Table 1: Treatment-Control Balance

	Control group (1)	Treatment group (2)	Difference T − C (3)	p-value (4)
Panel A. Grade 10				
Girl	0.534	0.524	−0.010	0.309
Non-repeater	0.802	0.801	−0.001	0.892
Non-French	0.059	0.061	0.002	0.652
High SES	0.374	0.381	0.007	0.364
Medium-high SES	0.136	0.129	−0.006	0.221
Medium-low SES	0.248	0.235	−0.012	0.064
Low SES	0.243	0.255	0.012	0.085
DNB percentile rank in math	45.08	44.95	−0.134	0.829
DNB percentile rank in French	45.95	46.32	0.365	0.555
Took at least one science elective course	0.391	0.396	0.005	0.820
Took at least one standard elective course	0.769	0.738	−0.031	0.138
N	6,801	6,899		
Test of joint significance	F-stat: 1.505 (p-value: 0.135)			
Panel B. Grade 12 (science track)				
Girl	0.498	0.485	−0.014	0.292
Non-repeater	0.745	0.767	0.021	0.023
Non-French	0.054	0.048	−0.006	0.275
High SES	0.441	0.470	0.029	0.008
Medium-high SES	0.144	0.143	−0.001	0.826
Medium-low SES	0.216	0.201	−0.015	0.023
Low SES (D)	0.198	0.186	−0.012	0.140
DNB percentile rank in math	62.04	62.42	0.387	0.548
DNB percentile rank in French	59.03	59.96	0.891	0.176
N	2,853	2,898		
Test of joint significance	F-stat: 1.565 (p-value: 0.138)			

Notes: Each row corresponds to a different model, based on administrative data on students. Columns 1 and 2 report the average values for students in the control and treatment groups respectively. Column 3 shows the coefficient on a treatment dummy in a regression of each variable on treatment, with p-values reported in column 4. The regression controls for school fixed effects to account for the fact that randomization was stratified by school. The F-statistic is from a test of the joint significance of the coefficients in a regression of the treatment group dummy on the baseline characteristics. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Student characteristics come from the *Bases Élèves académiques* of the three educational districts of Paris, Créteil and Versailles for the academic year 2015-2016. French and math scores are from the exams of the *Diplôme national du brevet* (DNB) in middle school.

Table 2: Female Facilitators: Summary Statistics

	All	Professionals (employed by firm)	Researchers (Ph.D./Postdoc)
Age	33.3	35.6	30.1
	(5.7)	(6.0)	(3.1)
Non-French	0.14	0.17	0.10
Has a Ph.D degree	0.60	0.39	0.95
Attended a Grande École	0.39	0.43	0.33
Field: Math, Physics, Engineering	0.23	0.14	0.38
Field: Earth and Life Sciences	0.64	0.66	0.62
Field: Other	0.13	0.20	0.00
Has children	0.42	0.56	0.19
Facilitator the year before	0.25	0.29	0.19
Number of high schools visited	1.8	1.6	2.1
	(0.8)	(0.7)	(0.9)
Number of classroom interventions	5.4	4.9	6.2
	(2.4)	(2.3)	(2.5)
N	56	35	21

Notes: The summary statistics are computed from the post-visit facilitator survey that was administered online to collect feedback about the classroom visits. Standard deviations are shown in parentheses.

Table 3: Perceptions of Science-Related Careers

	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)
Panel A. Grade 10						
Negative perceptions of science-related careers (index)	0.016	−0.245***	0.000	−0.032	−0.167***	0.000
		(0.028)			(0.029)	
Studies in science are lengthy	0.839	−0.087***	0.000	0.848	−0.074***	0.000
		(0.010)	[0.001]		(0.010)	[0.001]
Jobs in science are dreary	0.288	−0.032***	0.006	0.318	−0.006	0.633
		(0.012)	[0.011]		(0.013)	[0.634]
Jobs in science are solitary	0.325	−0.061***	0.000	0.298	−0.062***	0.000
		(0.012)	[0.001]		(0.011)	[0.001]
Higher wages in science	0.641	0.008	0.535	0.666	0.015	0.237
		(0.014)	[0.536]		(0.013)	[0.297]
Hard to maintain work-life balance	0.295	−0.026**	0.026	0.279	−0.029**	0.014
		(0.012)	[0.033]		(0.012)	[0.023]
N		6,475			5,751	
Panel B. Grade 12 (science track)						
Negative perceptions of science-related careers (index)	0.011	−0.312***	0.000	0.005	−0.155***	0.000
		(0.034)			(0.033)	
Studies in science are lengthy	0.669	−0.110***	0.000	0.724	−0.091***	0.000
		(0.015)	[0.001]		(0.014)	[0.001]
Jobs in science are dreary	0.171	−0.019	0.141	0.237	−0.026	0.114
		(0.013)	[0.141]		(0.016)	[0.143]
Jobs in science are solitary	0.233	−0.088***	0.000	0.206	−0.047***	0.000
		(0.012)	[0.001]		(0.013)	[0.001]
Higher wages in science	0.529	0.059***	0.001	0.574	0.027*	0.093
		(0.018)	[0.002]		(0.016)	[0.143]
Hard to maintain work-life balance	0.225	−0.049***	0.001	0.168	−0.012	0.260
		(0.015)	[0.002]		(0.011)	[0.260]
N		2,600			2,636	

Notes: This table reports treatment effects estimates on students' perceptions of science-related careers, separately by grade level and gender. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report both the standard p-value of the estimated treatment effect and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Perceptions of Gender Roles in Science

	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)
Panel A. Grade 10						
More men in science-related jobs	0.629	0.156***	0.000	0.630	0.168***	0.000
		(0.013)	[0.001]		(0.014)	[0.001]
Brains of women and men are different	0.212	−0.050***	0.000	0.208	−0.048***	0.000
		(0.010)	[0.001]		(0.011)	[0.001]
Men are more gifted in math	0.187	−0.026**	0.015	0.301	−0.048***	0.001
		(0.011)	[0.016]		(0.014)	[0.001]
Women like science less than men	0.162	0.059***	0.000	0.200	0.103***	0.000
		(0.011)	[0.001]		(0.013)	[0.001]
Women are discriminated in science careers	0.603	0.127***	0.000	0.525	0.153***	0.000
		(0.013)	[0.001]		(0.014)	[0.001]
N		6,475			5,751	
Panel B. Grade 12 (science track)						
More men in science-related jobs	0.714	0.125***	0.000	0.722	0.149***	0.000
		(0.016)	[0.001]		(0.015)	[0.001]
Brains of women and men are different	0.146	−0.023**	0.026	0.177	−0.038***	0.006
		(0.010)	[0.026]		(0.014)	[0.009]
Men are more gifted in math	0.160	−0.038***	0.002	0.265	−0.028*	0.072
		(0.012)	[0.003]		(0.015)	[0.073]
Women like science less than men	0.073	0.044***	0.000	0.148	0.073***	0.000
		(0.009)	[0.001]		(0.015)	[0.001]
Women are discriminated in science careers	0.625	0.095***	0.000	0.600	0.072***	0.000
		(0.020)	[0.001]		(0.018)	[0.001]
N		2,600			2,636	

Notes: This table reports treatment effects estimates on students' perceptions of gender roles in science, separately by grade level and gender. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 5 and 6 report both the standard p-value and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Stated Preferences and Self-Concept

	Girls			Boys		
	Control group mean (1)	Treatment effect (LATE) (2)	p-value (3)	Control group mean (4)	Treatment effect (LATE) (5)	p-value (6)
Panel A. Grade 10						
Taste for science subjects (index)	−0.191	−0.038	0.294	0.190	−0.019	0.533
		(0.036)			(0.031)	
Low math self-concept (index)	0.212	0.008	0.806	−0.235	−0.039	0.217
		(0.031)			(0.032)	
Interest in science-related careers (index)	−0.118	0.017	0.556	0.145	0.010	0.731
		(0.029)			(0.029)	
N		6,475			5,751	
Panel B. Grade 12 (science track)						
Taste for science subjects (index)	−0.005	0.016	0.632	0.000	0.000	0.998
		(0.034)			(0.039)	
Low math self-concept (index)	0.182	−0.050	0.202	−0.191	−0.072**	0.041
		(0.039)			(0.035)	
Interest in science-related careers (index)	−0.065	0.116***	0.002	0.067	0.050	0.126
		(0.037)			(0.033)	
N		2,600			2,636	

Notes: This table reports treatment effects estimates on students' interest in science, math self-concept, and interest in science-related careers, separately by grade level and gender. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report the p-value of the estimated treatment effect.

Table 6: Enrollment Status the Following Year

	Girls			Boys		
	Control group mean	Treatment effect (LATE)	p-value [q-value]	Control group mean	Treatment effect (LATE)	p-value [q-value]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A : Grade 10						
All STEM tracks						
Grade 11 : Science track	0.363	−0.008	0.586	0.576	−0.006	0.676
		(0.014)			(0.015)	
General vs. technical STEM track						
General STEM track	0.336	−0.002	0.888	0.432	0.004	0.773
		(0.014)	[0.889]		(0.014)	[0.774]
Technical STEM track	0.027	−0.006	0.112	0.145	−0.010	0.235
		(0.004)	[0.224]		(0.009)	[0.470]
N		7,241			6,459	
Panel B : Grade 12 (science track)						
All STEM majors						
Undergraduate major in STEM	0.282	0.024*	0.080	0.468	0.003	0.886
		(0.014)			(0.020)	
Selective vs. non-selective STEM						
Selective STEM major (CPGE)	0.109	0.035***	0.002	0.231	0.020	0.200
		(0.011)	[0.004]		(0.016)	[0.283]
Non-selective STEM major (University)	0.173	−0.011	0.322	0.237	−0.017	0.212
		(0.011)	[0.322]		(0.014)	[0.283]
Male- vs. non-male-dominated STEM						
Male-dominated STEM major (Math, Physics & Computer Science)	0.165	0.038***	0.002	0.377	0.017	0.387
		(0.012)	[0.004]		(0.019)	[0.388]
Non-male-dominated STEM major (Earth & Life Sciences)	0.118	−0.015	0.158	0.091	−0.014	0.119
		(0.010]	[0.211]		(0.009)	[0.283]
N		2,827			2,924	

Notes: This table reports treatment effects estimates on students' educational outcomes in the academic year following the classroom interventions, i.e., 2016-2017, separately by grade level and gender. Each row corresponds to a different model, based on administrative data on student enrollment outcomes. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report both the standard p-value and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Heterogenous Effects on Student Perceptions, by Level of Performance in Math

	Girls			Boys		
	Treatment effect (LATE) by level of performance in math			Treatment effect (LATE) by level of performance in math		
	Below median	Above median	Difference: p-value [q-value]	Below median	Above median	Difference: p-value [q-value]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Grade 10						
Studies in science are lengthy	−0.040***	−0.126***	0.000	−0.059***	−0.087***	0.199
	(0.015)	(0.014)	[0.001]	(0.015)	(0.015)	[0.499]
Men are more gifted in math	−0.011	−0.038***	0.167	−0.036*	−0.059***	0.410
	(0.017)	(0.012)	[0.279]	(0.022)	(0.018)	[0.513]
Women are discriminated in science careers	0.173***	0.088***	0.001	0.176***	0.134***	0.138
	(0.019)	(0.017)	[0.003]	(0.021)	(0.018)	[0.499]
Would consider a job in science	−0.012	0.001	0.628	0.011	0.038**	0.337
	(0.021)	(0.017)	[0.628]	(0.021)	(0.017)	[0.513]
Wages are important in choice of studies (z-score)	−0.048	0.019	0.259	0.005	0.009	0.938
	(0.042)	(0.041)	[0.324]	(0.042)	(0.034)	[0.939]
Grade 11: Science Track	−0.020	0.004	0.308	−0.019	0.009	0.330
	(0.015)	(0.018)	[0.308]	(0.021)	(0.018)	[0.330]
N	3,534	3,707		3,148	3,311	
Panel B. Grade 12 (science track)						
Studies in science are lengthy	−0.098***	−0.121***	0.587	−0.053**	−0.127***	0.049
	(0.026)	(0.025)	[0.587]	(0.022)	(0.025)	[0.123]
Men are more gifted in math	−0.024	−0.053***	0.325	0.012	−0.067***	0.028
	(0.019)	(0.020)	[0.587]	(0.024)	(0.023)	[0.123]
Women are discriminated in science careers	0.114***	0.075**	0.351	0.093***	0.052*	0.356
	(0.027)	(0.030)	[0.587]	(0.030)	(0.027)	[0.445]
Would consider a job in science	0.012	0.045*	0.380	0.001	0.053**	0.135
	(0.022)	(0.024)	[0.587]	(0.023)	(0.021)	[0.226]
Wages are important in choice of studies (z-score)	0.094*	0.147**	0.535	0.076	0.024	0.497
	(0.053)	(0.061)	[0.587]	(0.050)	(0.048)	[0.498]
College: Selective STEM track	0.001	0.066***	0.015	−0.012	0.040	0.127
	(0.013)	(0.020)	[0.030]	(0.018)	(0.027)	[0.255]
College: Major in male-dominated STEM	0.025	0.046**	0.499	−0.002	0.024	0.498
	(0.018)	(0.022)	[0.500]	(0.025)	(0.028)	[0.498]
N	1,544	1,482		1,497	1,544	

Notes: This table reports treatment effects estimates on students' perceptions and enrollment outcome the following year, separately by grade level, gender, and academic performance in math (above/below median). Each row corresponds to a different model, based on students' responses to the post-treatment survey and enrollment the following year. Student academic performance in math is measured from the grades they obtained on the final math exam of the *Diplôme national du Brevet* at the end of middle school for Grade 10 students, and from the grades obtained on the *Baccalauréat* math final exam for Grade 12 students. Columns 1 and 2 (for girls) and 4 and 5 (for boys) report the coefficients on the interactions between between a "facilitator visit" and dummies for the student being either below (columns 1 and 4) or above (columns 2 and 5) the median level of performance in math, in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report both the standard p-value of the difference between estimated treatment effects for students above and below the median performance in math, and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Heterogenous Effects on Student Perceptions, by Female Facilitator Background

	Girls			Boys		
	Treatment effect (LATE) by facilitator background			Treatment effect (LATE) by facilitator background		
	Researcher	Professional	Difference: p-value [q-value]	Researcher	Professional	Difference: p-value [q-value]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Grade 10						
Would consider a job in science	0.010	−0.007	0.499	0.011	0.030	0.408
	(0.024)	(0.022)	[0.659]	(0.022)	(0.019)	[0.545]
Higher wages in science	0.017	0.002	0.550	0.007	0.018	0.633
	(0.023)	(0.020)	[0.659]	(0.019)	(0.020)	[0.633]
Wages are important in choice of studies (z-score)	−0.001	−0.024	0.658	−0.011	0.032	0.331
	(0.045)	(0.044)	[0.659]	(0.041)	(0.037)	[0.545]
Hard to maintain work-life balance	−0.029	−0.020	0.654	−0.005	−0.045**	0.077
	(0.019)	(0.019)	[0.659]	(0.022)	(0.017)	[0.307]
Grade 11: Science Track	0.000	−0.001	0.975	−0.025	0.009	0.218
	(0.025)	(0.026)	[0.975]	(0.026)	(0.024)	[0.219]
N	3,237	4,004		2,890	3,569	
Panel B. Grade 12 (science track)						
Would consider jobs in science	−0.027	0.074***	0.001	0.008	0.042*	0.242
	(0.028)	(0.025)	[0.006]	(0.027)	(0.023)	[0.485]
Higher wages in science	0.020	0.081***	0.091	0.039	0.020	0.594
	(0.035)	(0.027)	[0.122]	(0.032)	(0.029)	[0.595]
Wages are important in choice of studies (z-score)	0.041	0.163***	0.081	0.029	0.068	0.529
	(0.071)	(0.055)	[0.122]	(0.060)	(0.051)	[0.595]
Hard to maintain work-life balance	−0.049*	−0.046*	0.931	0.003	−0.029	0.158
	(0.029)	(0.026)	[0.932]	(0.022)	(0.019)	[0.485]
College: Selective STEM track	0.001	0.052**	0.048	0.002	0.033	0.424
	(0.022)	(0.022)	[0.097]	(0.033)	(0.033)	[0.425]
College: Major in male-dominated STEM	0.024	0.042*	0.546	−0.023	0.047	0.087
	(0.025)	(0.022)	[0.546]	(0.037)	(0.032)	[0.174]
N	1,180	1,647		1,312	1,612	

Notes: This table reports treatment effects estimates on students' perceptions, separately by grade level, gender, and by background of facilitator (professional or researcher) who visited the classroom. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 2 (for girls) and 4 and 5 (for boys) report the coefficients on the interactions between between a "facilitator visit" and dummies for the facilitator being either a researcher (columns 1 and 4) or a professional (columns 2 and 5), in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report the standard p-value for the difference in the treatment effects estimates for students above vs. below the median performance in math. The p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes are reported in brackets, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(For Online Publication)

Appendix to

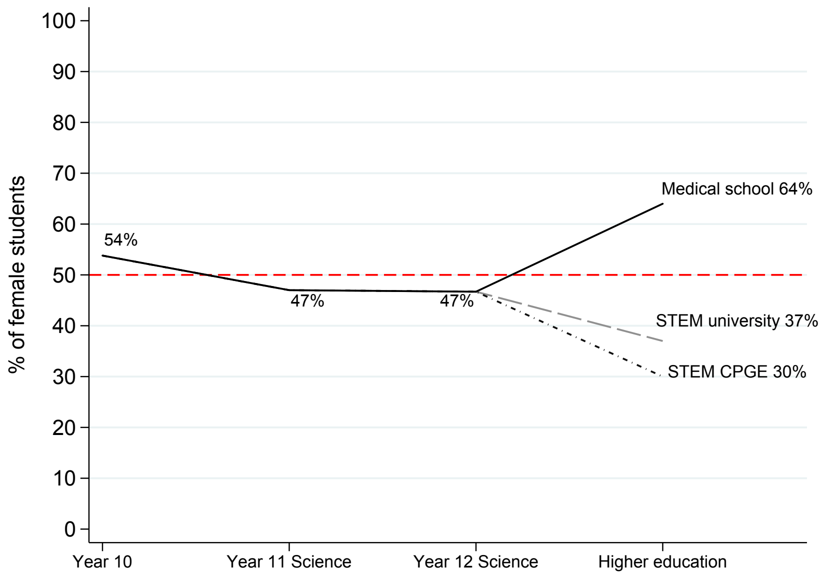
Can female Role Models Reduce the Gender Gap in Science? Evidence from Classroom Interventions in French High Schools

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

December 2018

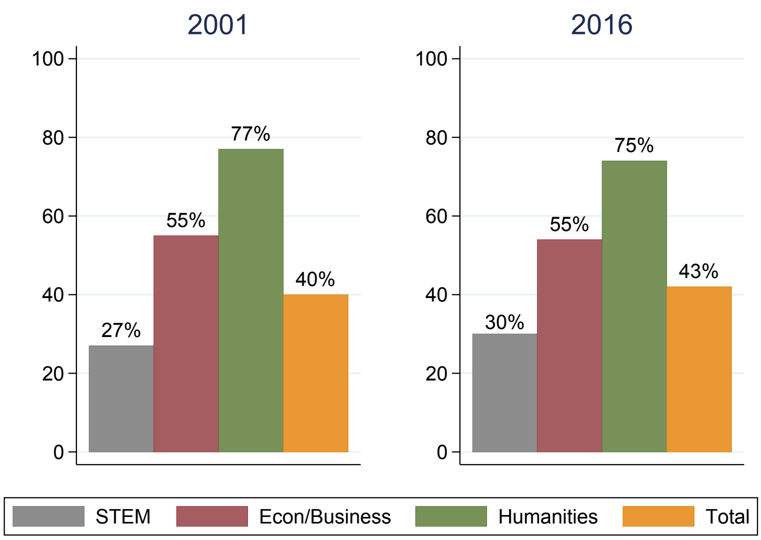
List of Appendices

Appendix A Female Representation in STEM Studies in France



Notes: The figure shows the share of female students in scientific majors in high school and in higher education for the whole French population of students in 2016-2017.
Source: MENESR (2018).

Figure A1: Female Underrepresentation in STEM Studies, France, 2016-2017



Notes: The figure shows the share of female students in selective two-year undergraduate academic programs (*Classes Préparatoires aux Grandes Écoles*) for the whole French population of students in 2001-2002 and 2016-2017.
Source: MENESR (2018).

Figure A2: Share of Female Students in Selective Undergraduate Programs (CPGE), France

Appendix B Program Details

B.1 Program Timeline

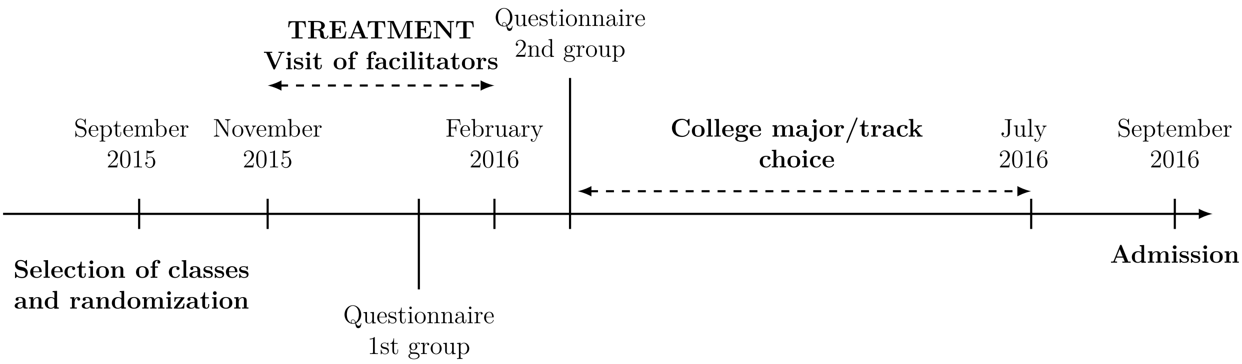
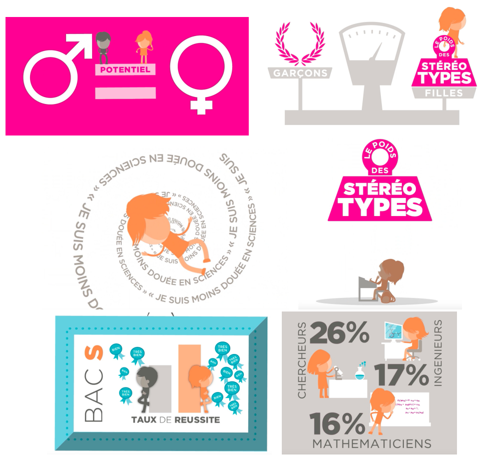


Figure B3: Program Timeline

B.2 Videos Shown During Classroom Interventions

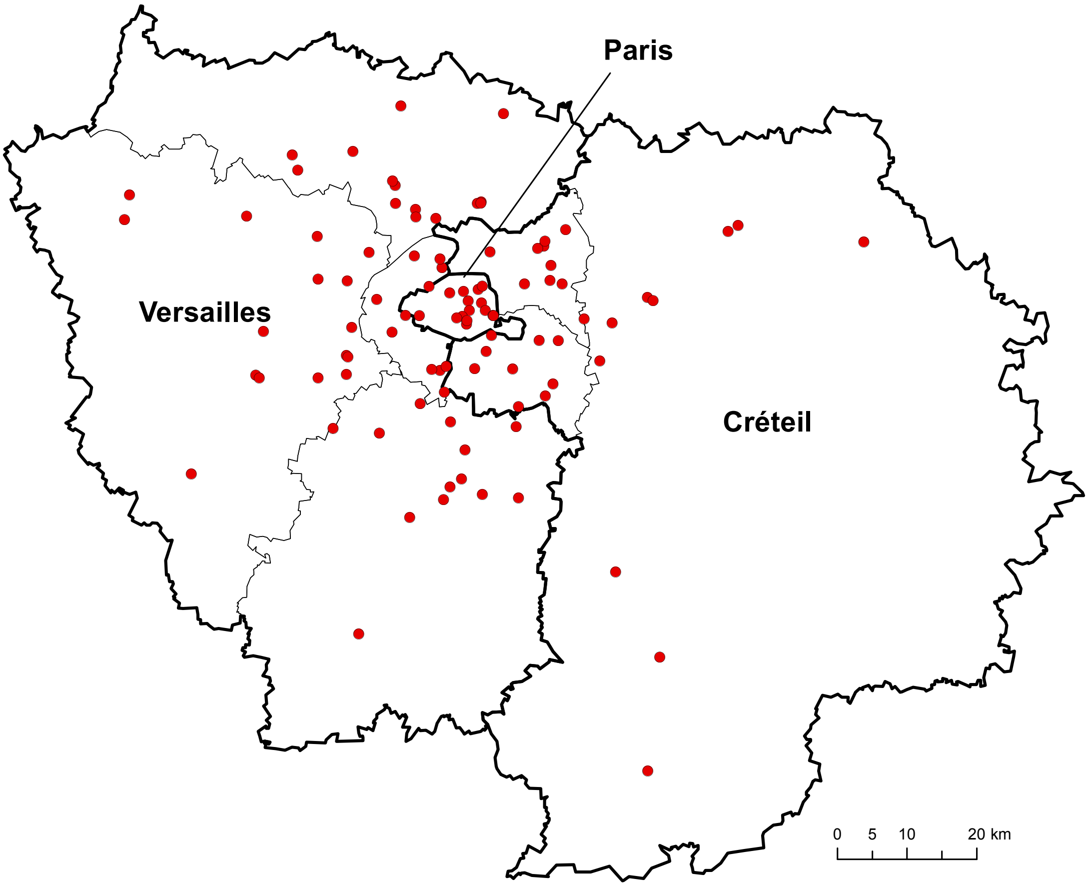
(a) Video 1: "Jobs in Science: Beliefs or Reality?"



(b) Video 2: "Are we All Equal in Science?"

Figure B4: Thumbnails of Videos Shown During Classroom Interventions

B.3 Experimental Sample



Notes: This map represents the three educational districts (*académies*) of the Paris region: Paris, Créteil and Versailles. The solid purple dots show the location of the 98 high schools that participated in the program.

Figure B5: Participating High Schools

Table B1: Experimental Sample - Summary Statistics

	High schools operating in the Paris region	Participating high schools	
		Classes selected for random assignment	Classes not selected for random assignment
	(1)	(2)	(3)
Number of high schools	489	98	96
Share private	0.22	0.17	0.08
Panel A. Grade 10			
Number of students	115,720	13,700	19,147
Number of classes	3,627	416	592
Female	0.53	0.53	0.52
Age (years)	15.14	15.13	15.14
High SES	0.40	0.38	0.36
Medium-high SES	0.12	0.13	0.13
Medium-low SES	0.24	0.24	0.25
Low SES	0.24	0.25	0.26
DNB percentile rank in math	57.69	58.48	55.10
DNB percentile rank in French	57.23	57.85	55.75
Panel B. Grade 12 (science track)			
Number of students	38,582	5,751	5,623
Number of classes	1,267	185	179
Female	0.46	0.49	0.42
Age (years)	17.11	17.12	17.10
High SES	0.52	0.46	0.53
Medium-high SES	0.12	0.14	0.13
Medium-low SES	0.20	0.21	0.18
Low SES	0.16	0.19	0.16
DNB percentile rank in math	76.25	74.06	76.20
DNB percentile rank in French	70.78	69.61	69.78

Notes: This table compares the characteristics of high schools that participated in the program to those of all high schools operating in the Paris region. Among participating schools, Grade 10 and Grade 12 (science track) classes that were selected by principals for random assignment to treatment are compared to classes that were not selected. The summary statistics are computed from the *Bases Élèves académiques* of the three educational districts of Paris, Créteil and Versailles for the academic year 2015-2016. French and math scores are from the exams of the *Diplôme national du brevet* (DNB) in middle school.

Appendix C Compliance with Randomization

Table C2: Compliance with Randomization

	All classrooms (1)	Classrooms assigned to Control group (2)	Classrooms assigned to Treatment group (3)
Panel A. Grade 10			
Number of classes visited by female facilitator	199	2	197
Number of classes not visited by female facilitator	217	205	12
Number of students	13,700	6,801	6,899
Student-level compliance with random assignment	0.97	0.99	0.94
Panel B. Grade 12 (science track)			
Number of classes visited by female facilitator	91	2	89
Number of classes not visited by female facilitator	94	90	4
Number of students	5,751	2,853	2,898
Student-level compliance with random assignment	0.97	0.98	0.95

Notes: This table reports compliance with the random assignment of Grade 10 and Grade 12 (science track) classes to the treatment and control groups. Two-way non compliance was due to either classes in the treatment not being visited by the facilitators or to classes in the control group being visited by facilitators.

Appendix D Student Survey Non-Response

D.1 Student Survey Response Rates

Table D3: Student Post-Treatment Survey – Response Rates

	Control group (1)	Treatment group (2)	Difference T − C (3)	Difference: p-value (4)
Panel A. Grade 10				
Survey response rate	0.879	0.905	0.026	0.026
	(0.009)	(0.004)	(0.012)	
Number of students	6,801	6,899		
Panel B. Grade 12 (science track)				
Survey response rate	0.908	0.913	0.005	0.693
	(0.009)	(0.005)	(0.012)	
Number of students	2,853	2,898		

Notes: This table reports the student survey response rate for students in the Grade 10 and Grade 12 classes that participated in the program. The response rates are computed with respect to the list of all students who were recorded in the *Bases Élèves académiques* as being enrolled in these classes during the academic year 2015-2016. The regression controls for school fixed effects to account for the fact that randomization was stratified by school.

D.2 Student Survey: Characteristics of Respondents

Table D4: Treatment-Control Balance – Survey Respondents

	Control group (1)	Treatment group (2)	Difference T − C (3)	Difference: p-value (4)
Panel A. Grade 10				
Girl	0.537	0.523	−0.014	0.160
Non-repeater	0.809	0.809	0.000	0.983
Non-French	0.057	0.060	0.003	0.528
High SES	0.380	0.384	0.004	0.622
Medium-high SES	0.136	0.132	−0.005	0.392
Medium-low SES	0.245	0.236	−0.009	0.200
Low SES	0.239	0.249	0.010	0.158
DNB percentile rank in math	45.58	45.68	0.102	0.873
DNB percentile rank in French	46.43	46.79	0.357	0.578
Took at least one science elective course	0.394	0.403	0.009	0.693
Took at least one standard elective course	0.772	0.739	−0.032	0.132
N	5,981	6,245		
Test of joint significance	F-stat: 1.147 (p-value: 0.326)			
Panel B. Grade 12 (science track)				
Girl	0.504	0.490	−0.014	0.319
Non-repeater	0.755	0.775	0.020	0.038
Non-French	0.053	0.045	−0.008	0.129
High SES	0.436	0.476	0.039	0.001
Medium-high SES	0.147	0.145	−0.001	0.852
Medium-low SES	0.219	0.196	−0.022	0.001
Low SES	0.198	0.183	−0.016	0.086
DNB percentile rank in math	62.43	62.52	0.089	0.899
DNB percentile rank in French	59.35	60.03	0.673	0.323
N	2,594	2,642		
Test of joint significance	F-stat: 2.445 (p-value: 0.016)			

Notes: Each row corresponds to a different model, based on administrative data on students. The sample is restricted to students in the treatment and control groups who answered the post-treatment survey. Columns 1 and 2 report the average values for students in the control and treatment groups respectively. Column 3 shows the coefficient on a treatment dummy in a regression of each variable on treatment, with p-values reported in column 4. The regression controls for school fixed effects to account for the fact that randomization was stratified by school. The F-statistic is from a test of the joint significance of the coefficients in a regression of the treatment group dummy on the baseline characteristics. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Student characteristics come from the *Bases Élèves académiques* of the three educational districts of Paris, Créteil and Versailles for the academic year 2015-2016. French and math scores are from the exams of the *Diplôme national du brevet* (DNB) in middle school.

Appendix E Post-Visit Facilitator Survey

Table E5: Post-Visit Facilitator Survey – Summary Statistics

	Facilitator background			Difference R−P	p-value of difference
	All (1)	Professionals (2)	Researchers (3)	(4)	(5)
A. Adults present during the intervention					
Teacher was present	0.890	0.896	0.883	−0.013	0.773
Teacher's subject: Science	0.589	0.589	0.589	0.000	0.997
Teacher's gender: female	0.558	0.568	0.545	−0.023	0.682
Teacher was interested	0.692	0.731	0.642	−0.089	0.114
Other adult present beside teacher	0.348	0.316	0.390	0.074	0.237
B. General atmosphere during the intervention					
Students engaged in the discussion	0.386	0.391	0.378	−0.013	0.838
Students were inattentive	0.134	0.111	0.164	0.053	0.260
Students were very interested	0.423	0.422	0.425	0.003	0.963
Powerpoint worked well	0.963	0.982	0.938	−0.044	0.172
Videos worked well	0.888	0.886	0.891	0.004	0.940
Organizational problems	0.160	0.141	0.184	0.043	0.488
Talk interrupted due to discipline problems	0.068	0.060	0.078	0.018	0.651
C. Students' responsiveness to topics addressed during the intervention					
Responsive to 'science is everywhere'	0.430	0.469	0.380	−0.089	0.360
Responsive to 'jobs in science are fulfilling'	0.352	0.314	0.400	0.086	0.332
Responsive to 'girls can work in science too'	0.375	0.391	0.355	−0.036	0.675
Responsive to 'jobs in science pay'	0.387	0.474	0.275	−0.199	0.046
Responsive to the short videos	0.546	0.589	0.490	−0.099	0.340
D. Overall impression of the facilitator					
Gender stereotypes were strong	0.402	0.467	0.318	−0.150	0.075
Overall feedback was positive	0.556	0.572	0.536	−0.036	0.669
Interventin was well designed	0.474	0.494	0.450	−0.044	0.661
Number of classroom interventions	301	170	131		

Notes: The summary statistics are computed from the post-visit facilitator survey that was administered online to collect feedback about the classroom visits.

Appendix F Classrooms Assigned to Professionals and Researchers

Table F6: Balancing Test – Classrooms Assigned to Professionals and Researchers

	Classrooms assigned to		Difference R−P	p-value of difference
	Professionals	Researchers		
	(1)	(2)	(3)	(4)
Panel A. Grade 10				
Girl	0.529	0.528	0.000	0.976
Non-repeater	0.797	0.808	0.011	0.184
Non-French	0.057	0.064	0.007	0.259
High SES	0.405	0.343	−0.063	0.002
Medium-high SES	0.128	0.137	0.009	0.235
Medium-low SES	0.235	0.249	0.014	0.160
Low SES	0.231	0.271	0.040	0.018
DNB percentile rank in math	45.69	44.18	−1.511	0.328
DNB percentile rank in French	46.95	45.13	−1.818	0.183
Took at least one science elective course	0.380	0.412	0.032	0.357
Took at least one standard elective course	0.741	0.768	0.028	0.291
N	7,573	6,127		
Test of joint significance	F-stat: 0.735 (p-value: 0.691)			
Panel B. Grade 12 (science track)				
Girl	0.505	0.474	−0.032	0.114
Non-repeater	0.748	0.766	0.018	0.267
Non-French	0.046	0.057	0.010	0.272
High SES	0.474	0.431	−0.043	0.196
Medium-high SES	0.137	0.152	0.016	0.216
Medium-low SES	0.205	0.213	0.009	0.544
Low SES	0.184	0.203	0.019	0.428
DNB percentile rank in math	63.34	60.78	−2.562	0.185
DNB percentile rank in French	61.05	57.44	−3.615	0.035
N	3,259	2,492		
Test of joint significance	F-stat: 0.489 (p-value: 0.863)			

Notes: Columns 1 and 2 report the average values for students in the classrooms that were randomly selected to be visited by a professional or by a researcher, respectively. Column 3 shows the coefficient on a dummy equal to one if the student received the visit of a researcher, with p-values reported in column 4. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Student characteristics come from the *Bases Élèves académiques* of the three educational districts of Paris, Créteil and Versailles for the academic year 2015-2016. French and math scores are from the exams of the *Diplôme national du brevet* (DNB) in middle school.

Appendix G Effects on Students' Preferences and Self-Concept

G.1 Taste for Science Subjects

Table G7: Taste for Science Subjects

	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)
Panel A. Grade 10						
Taste for science subjects (index)	−0.191	−0.038	0.294	0.190	−0.019	0.533
		(0.036)			(0.031)	
Enjoys math (z-score)	−0.161	−0.002	0.961	0.186	−0.002	0.935
		(0.034)	[0.961]		(0.031)	[0.935]
Enjoys physics-chemistry (z-score)	−0.189	−0.040	0.289	0.218	−0.022	0.505
		(0.038)	[0.578]		(0.033)	[0.935]
Enjoys biology-geoscience (z-score)	−0.064	−0.058	0.137	0.074	−0.027	0.443
		(0.039)	[0.548]		(0.035)	[0.935]
Enjoys science: Agree	0.657	−0.011	0.444	0.789	0.003	0.804
		(0.015)	[0.593]		(0.012)	[0.935]
N		6,475			5,751	
Panel B. Grade 12 (science track)						
Taste for science subjects (index)	−0.005	0.016	0.632	0.000	0.000	0.998
		(0.034)			(0.039)	
Enjoys math (z-score)	−0.101	0.067*	0.089	0.102	0.075*	0.063
		(0.040)	[0.357]		(0.040)	[0.203]
Enjoys physics-chemistry (z-score)	−0.088	−0.001	0.984	0.090	−0.021	0.598
		(0.044)	[0.984]		(0.040)	[0.599]
Enjoys biology-geoscience (z-score)	0.199	−0.030	0.435	−0.207	−0.059	0.318
		(0.038)	[0.871]		(0.059)	[0.424]
Enjoys science: Agree	0.918	−0.001	0.887	0.929	0.013	0.101
		(0.009)	[0.984]		(0.008)	[0.203]
N		2,600			2,636	

Notes: This table reports treatment effects estimates on students' taste for science subjects taught at school, separately by grade level and gender. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report both the standard p-value and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

G.2 Interest in Science-Related Careers

Table G8: Interest in Science-Related Careers

	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)
Panel A. Grade 10						
Interest in science-related careers (index)	−0.118	0.017 (0.029)	0.556	0.145	0.010 (0.029)	0.731
Some jobs in science are interesting	0.844	0.019** (0.009)	0.050 [0.200]	0.855	0.000 (0.010)	1.000 [1.000]
Would consider a job in science	0.463	−0.004 (0.015)	0.776 [0.776]	0.588	0.023* (0.014)	0.089 [0.179]
Interested in at least one of listed STEM job[a]	0.496	0.022 (0.014)	0.105 [0.210]	0.805	0.021* (0.011)	0.052 [0.179]
Wages are important in choice of studies (z-score)	−0.041	−0.012 (0.029)	0.682 [0.776]	0.047	0.007 (0.027)	0.792 [1.000]
N		6,475			5,751	
Panel B. Grade 12 (science track)						
Interest in science-related careers (index)	−0.065	0.116*** (0.037)	0.002	0.067	0.050 (0.033)	0.126
Some jobs in science are interesting	0.961	0.013** (0.005)	0.013 [0.026]	0.938	0.021*** (0.008)	0.005 [0.022]
Would consider a job in science	0.716	0.031** (0.013)	0.019 [0.026]	0.764	0.030** (0.014)	0.029 [0.058]
Interested in at least one of listed STEM job[a]	0.638	0.012 (0.015)	0.415 [0.415]	0.844	0.006 (0.012)	0.634 [0.634]
Wages are important in choice of studies (z-score)	−0.040	0.119*** (0.038)	0.002 [0.007]	0.041	0.049 (0.031)	0.111 [0.149]
N		2,600			2,636	

Notes: This table reports treatment effects estimates on students' self-reported interest in science-related careers, separately by grade level and gender. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report both the standard p-value and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. [a]: the listed STEM jobs were computer scientist, engineer, renewable energies technician, and industrial designer.

G.3 Math Self-Concept

Table G9: Math Self-Concept

	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)
Panel A. Grade 10						
Low math self-concept (index)	0.212	0.008 (0.031)	0.806	−0.235	−0.039 (0.032)	0.217
Self-assessed performance in math (z-score)	−0.145	−0.016 (0.034)	0.634 [0.634]	0.167	0.021 (0.032)	0.502 [0.642]
Lost in front of a math problem	0.557	0.010 (0.014)	0.478 [0.634]	0.343	−0.007 (0.013)	0.610 [0.642]
Worried when thinking about math	0.622	−0.025* (0.013)	0.052 [0.109]	0.416	−0.032** (0.015)	0.028 [0.111]
Can succeed in science subjects if works hard enough	0.842	0.018* (0.009)	0.054 [0.109]	0.884	−0.004 (0.008)	0.642 [0.642]
N		6,475			5,751	
Panel B. Grade 12 (science track)						
Low math self-concept (index)	0.182	−0.050 (0.039)	0.202	−0.191	−0.072** (0.035)	0.041
Self-assessed performance in math (z-score)	−0.127	0.039 (0.038)	0.304 [0.406]	0.130	0.079** (0.038)	0.038 [0.077]
Lost in front of a math problem	0.488	−0.028 (0.020)	0.168 [0.336]	0.325	−0.028* (0.016)	0.072 [0.096]
Worried when thinking about math	0.556	−0.037** (0.019)	0.048 [0.193]	0.379	−0.051*** (0.016)	0.002 [0.007]
Can succeed in science subjects if works hard enough	0.941	−0.005 (0.007)	0.512 [0.512]	0.951	0.006 (0.007)	0.384 [0.385]
N		2,600			2,636	

Notes: This table reports treatment effects estimates on students' math self-concept, separately by grade level and gender. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report both the standard p-value and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

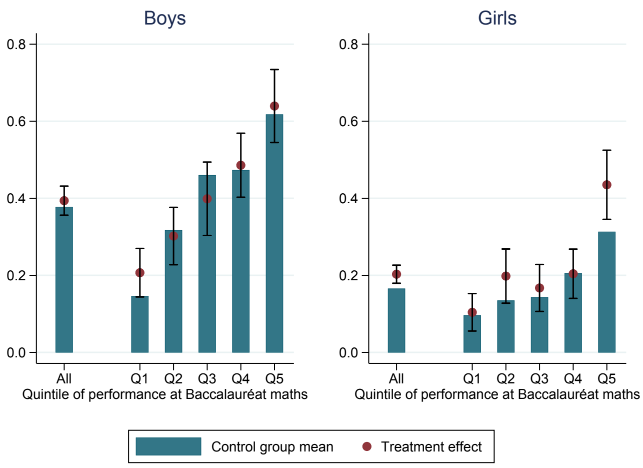
G.4 Main Survey Outcomes – with Baseline Controls

Table G10: Main Survey Outcomes - With Baseline Controls

	Girls			Boys		
	Control group mean	Treatment effect (LATE)	p-value [q-value]	Control group mean	Treatment effect (LATE)	p-value [q-value]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Grade 10						
Negative perceptions of science-related careers (index)	0.016	−0.242*** (0.027)	0.000	−0.032	−0.160*** (0.027)	0.000
More men in science-related jobs	0.629	0.154*** (0.013)	0.000 [0.001]	0.630	0.170*** (0.014)	0.000 [0.001]
Brains of women and men are different	0.212	−0.048*** (0.010)	0.000 [0.001]	0.208	−0.044*** (0.011)	0.000 [0.001]
Men are more gifted in math	0.187	−0.027*** (0.010)	0.009 [0.010]	0.301	−0.048*** (0.014)	0.001 [0.001]
Women like science less than men	0.162	0.056*** (0.011)	0.000 [0.001]	0.200	0.101*** (0.013)	0.000 [0.001]
Women are discriminated in science careers	0.603	0.125*** (0.013)	0.000 [0.001]	0.525	0.154*** (0.014)	0.000 [0.001]
Taste for science subjects (index)	−0.191	−0.040 (0.031)	0.191	0.190	−0.024 (0.026)	0.357
Low math self-concept (index)	0.212	0.009 (0.028)	0.760	−0.235	−0.030 (0.028)	0.288
Interest in science-related careers (index)	−0.118	0.008 (0.029)	0.792	0.145	0.005 (0.027)	0.862
N		6,475			5,751	
Panel B. Grade 12 (science track)						
Negative perceptions of science-related careers (index)	0.011	−0.297*** (0.032)	0.000	0.005	−0.169*** (0.033)	0.000
More men in science-related jobs	0.714	0.123*** (0.016)	0.000 [0.001]	0.722	0.147*** (0.015)	0.000 [0.001]
Brains of women and men are different	0.146	−0.025** (0.010)	0.014 [0.018]	0.177	−0.034** (0.014)	0.018 [0.023]
Men are more gifted in math	0.160	−0.028** (0.012)	0.023 [0.023]	0.265	−0.032** (0.016)	0.046 [0.047]
Women like science less than men	0.073	0.042*** (0.009)	0.000 [0.001]	0.148	0.073*** (0.015)	0.000 [0.001]
Women are discriminated in science	0.625	0.085*** (0.019)	0.000 [0.001]	0.600	0.076*** (0.018)	0.000 [0.001]
Taste for science subjects (index)	−0.005	0.017 (0.033)	0.600	0.000	0.012 (0.040)	0.756
Low math self-concept (index)	0.182	−0.055 (0.035)	0.109	−0.191	−0.065** (0.033)	0.048
Interest in science-related careers (index)	−0.065	0.107*** (0.036)	0.003	0.067	0.064* (0.035)	0.069
N		2,600			2,636	

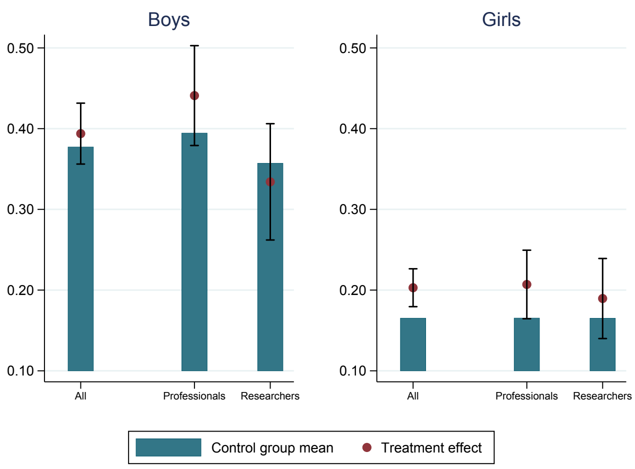
Notes: This table reports treatment effects estimates on students' perceptions of science-related careers and of gender roles in science, on stated preferences and math self-concept, separately by grade level and gender controlling for baseline characteristics used in the balancing checks. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report both the standard p-value and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix H Effects on Students' Enrollment the Following Year



Notes: The figure shows the fraction of Grade 12 (science track) students who enrolled in male-dominated STEM undergraduate programs after graduating for high school, for boys (left panel) and girls (right panel) separately. The blue bars indicate the overall mean among students in the control group and the separate means by quintile of *Baccalauréat* performance in math. The red solid dots show the estimated treatment effects with 95 percent confidence intervals denoted by vertical capped bars.

Figure H6: Grade 12 (Science Track) – Enrollment in Male-Dominated STEM the Following Year, by Student Gender and Quintile of *Baccalauréat* Performance in Math



Notes: The figure shows the fraction of Grade 12 (science track) students who enrolled in male-dominated STEM undergraduate programs after graduating for high school, for boys (left panel) and girls (right panel) separately. The blue bars indicate the overall mean among students in the control group and the separate means by background of the female facilitator who visited the classroom (professional or researcher). The red solid dots show the estimated treatment effects with 95 percent confidence intervals denoted by vertical capped bars.

Figure H7: Grade 12 (Science Track) – Enrollment in Male-Dominated STEM the Following Year, by Student Gender and Female Facilitator Background

Appendix I Persistence of Effects

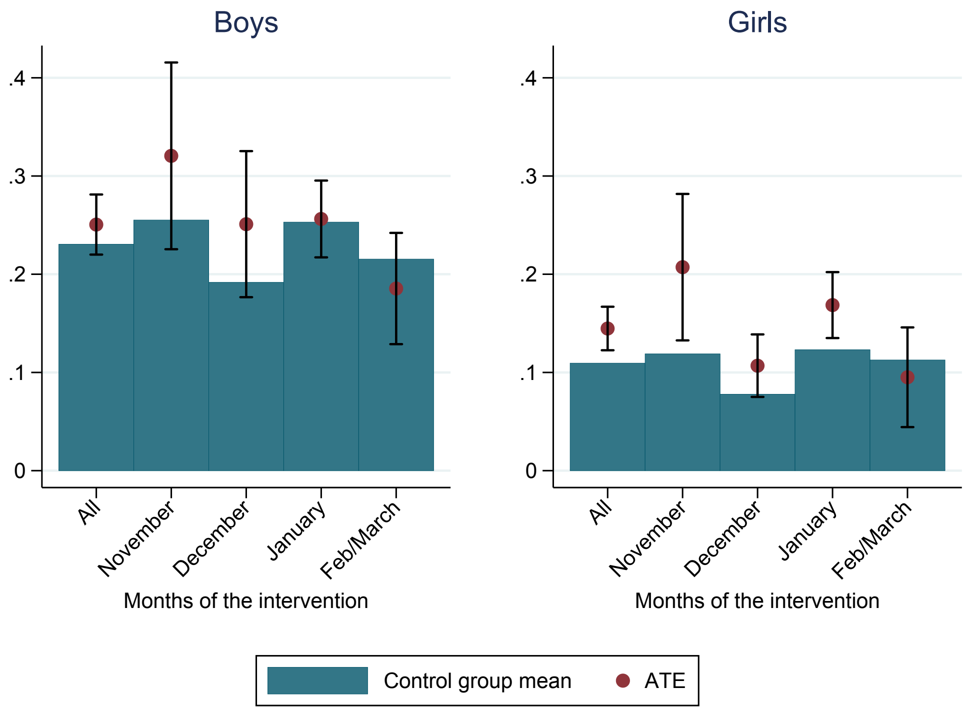
Table I11: Persistence of Effects on Student Perceptions

	Girls			Boys		
	Time elaspsed since intervention			Time elaspsed since intervention		
	1 to 2 months (1)	3 to 4 months (2)	5 to 6 months (3)	1 to 2 months (4)	3 to 4 months (5)	5 to 6 months (6)
Panel A : Grade 10						
Negative perceptions of science-related careers (index)	−0.398***	−0.200***	−0.143*	−0.194***	−0.168***	−0.049
	(0.053)	(0.037)	(0.077)	(0.055)	(0.036)	(0.083)
Men are more gifted in math: True	−0.041**	−0.035**	0.041*	−0.043	−0.047***	−0.098***
	(0.019)	(0.014)	(0.021)	(0.027)	(0.018)	(0.029)
Brains of women and men are different: True	−0.051***	−0.042***	−0.077***	−0.109***	−0.031**	0.003
	(0.018)	(0.013)	(0.027)	(0.017)	(0.013)	(0.030)
Women like science less than men: True	0.050**	0.067***	0.041	0.131***	0.107***	0.017
	(0.021)	(0.014)	(0.026)	(0.020)	(0.016)	(0.040)
More men in science-related jobs: True	0.164***	0.154***	0.164***	0.208***	0.163***	0.116***
	(0.021)	(0.017)	(0.033)	(0.023)	(0.018)	(0.039)
Women are discriminated in science: True	0.146***	0.135***	0.081**	0.161***	0.174***	0.110***
	(0.021)	(0.017)	(0.039)	(0.026)	(0.017)	(0.036)
Taste for science subjects (index)	0.039	−0.035	−0.053	0.012	−0.008	0.043
	(0.072)	(0.043)	(0.075)	(0.057)	(0.041)	(0.072)
Low math self-concept (index)	0.034	−0.006	−0.044	0.046	−0.103***	−0.088
	(0.055)	(0.039)	(0.080)	(0.062)	(0.039)	(0.090)
Interest in science-related careers (index)	0.085	0.003	0.021	−0.013	0.025	0.005
	(0.059)	(0.036)	(0.062)	(0.050)	(0.038)	(0.072)
N	1,993	3,716	831	1,803	3,318	693
Panel B : Grade 12 (science track)						
Negative perceptions of science-related careers (index)	−0.388***	−0.253***	−0.353***	−0.161***	−0.169***	−0.003
	(0.055)	(0.043)	(0.118)	(0.059)	(0.044)	(0.095)
Men are more gifted in math: True	−0.020	−0.058***	−0.033	−0.038	−0.028	0.005
	(0.023)	(0.016)	(0.040)	(0.030)	(0.018)	(0.060)
Brains of women and men are different: True	−0.041**	−0.022*	0.013	−0.064**	−0.036*	0.013
	(0.019)	(0.012)	(0.040)	(0.025)	(0.019)	(0.027)
Women like science less than men: True	0.050***	0.040***	0.032*	0.044	0.077***	0.144***
	(0.019)	(0.011)	(0.018)	(0.027)	(0.019)	(0.032)
More men in science-related jobs: True	0.136***	0.107***	0.208***	0.120***	0.159***	0.208***
	(0.029)	(0.019)	(0.060)	(0.024)	(0.021)	(0.046)
Women are discriminated in science: True	0.087**	0.102***	0.087	0.089***	0.085***	−0.011
	(0.035)	(0.023)	(0.072)	(0.025)	(0.024)	(0.062)
Taste for science subjects (index)	−0.035	−0.028	0.258***	0.007	0.010	−0.090
	(0.064)	(0.045)	(0.060)	(0.072)	(0.049)	(0.111)
Low math self-concept (index)	−0.065	−0.001	−0.169	0.016	−0.114**	−0.126
	(0.057)	(0.053)	(0.122)	(0.053)	(0.046)	(0.149)
Interest in science-related careers (index)	0.042	0.128***	0.255***	0.034	0.050	0.100
	(0.078)	(0.043)	(0.077)	(0.043)	(0.046)	(0.117)
N	801	1,468	394	805	1,514	370

Notes: This table reports treatment effects estimates on students' perceptions, separately by grade level, gender, and by time elapsed since the classroom interevention. Each row corresponds to a different model, based on students' responses to the post-treatment survey. For different time intervals between the classroom visit and the post-treatment student survey, the columns report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix J Heterogeneity by Timing of Visits

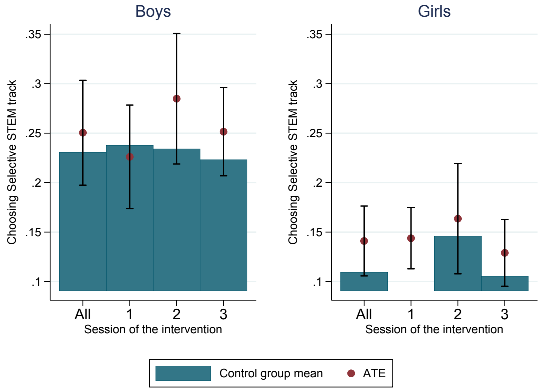
J.1 Enrollment Outcomes by Month of Intervention – Grade 12



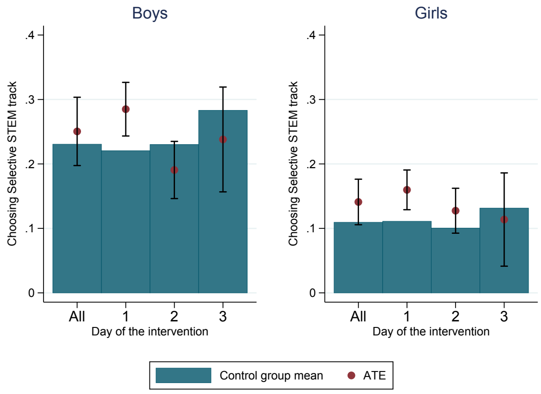
Notes: The figure shows the fraction of Grade 12 (science track) students who enrolled in selective STEM undergraduate programs after graduating for high school, for boys (left panel) and girls (right panel) separately. The blue bars indicate the overall mean among students in the control group and the separate means by month of classroom visit. The red solid dots show the estimated treatment effects with 95 percent confidence intervals denoted by vertical capped bars.

Figure J8: Enrollment in Selective STEM Undergraduate Program by Month of Intervention – Grade 12 (Science Track)

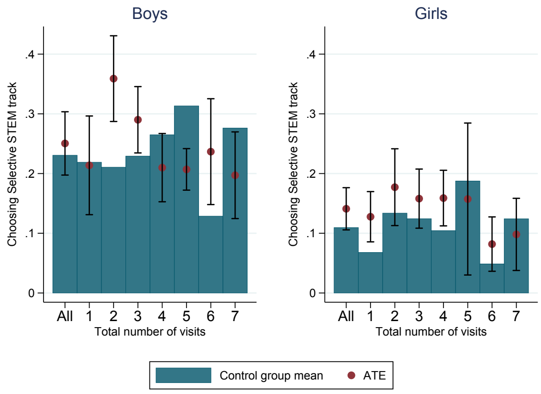
J.2 Enrollment outcomes by Order of Visit – Grade 12



(a) By Order of Classroom Visit in School



(b) By Order of School Visit



(c) By Overall Order of Classroom Visit

Notes: The figure shows the fraction of Grade 12 (science track) students who enrolled in selective STEM undergraduate programs after graduating for high school, for boys (left panel) and girls (right panel) separately. The blue bars indicate the overall mean among students in the control group and the separate means (i) by order of classroom visit in school (panel a), (ii) by order of school visit (panel b), and (iii) by overall order of classroom visit (panel c). The red solid dots show the estimated treatment effects with 95 percent confidence intervals denoted by vertical capped bars.

Figure J9: Impact of Interventions on Enrollment in Selective STEM Undergraduate Programs, by Order of Intervention, Grade 12 (science track)

Appendix K Spillover Effects

Table K12: Summary Statistics on Spillovers

	All (1)	Girls (2)	Boys (3)	p-value (4)
Panel A. Grade 10				
Control Group				
Was exposed to a science awareness program this year	0.146	0.148	0.144	0.655
Has been exposed to a science awareness program in the past	0.322	0.333	0.309	0.049
Never heard of a classroom visit in other classes	0.859	0.853	0.866	0.147
Vaguely heard of a classroom visit in other classes	0.122	0.127	0.117	0.216
Definitely heard of a classroom visit in other classes	0.018	0.020	0.017	0.462
Treatment Group				
Discussed the classroom visit				
with classmates	0.580	0.656	0.498	0.000
with other students from school	0.240	0.277	0.200	0.000
with other students outside of school	0.203	0.247	0.155	0.000
Was exposed to another science awareness program this year	0.128	0.120	0.138	0.041
Has been exposed to other science awareness programs in the past	0.182	0.149	0.218	0.000
Panel B. Grade 12 (science track)				
Control Group				
Was exposed to a science awareness program this year	0.287	0.284	0.291	0.694
Has been exposed to a science awareness program in the past	0.488	0.514	0.461	0.007
Never heard of a classroom visit in other classes	0.661	0.646	0.676	0.116
Vaguely heard of a classroom visit in other classes	0.292	0.308	0.275	0.063
Definitely heard of a classroom visit in other classes	0.047	0.045	0.049	0.634
Treatment Group				
Discussed the classroom visit				
with classmates	0.629	0.705	0.556	0.000
with other students from school	0.269	0.334	0.206	0.000
with other students outside of school	0.202	0.275	0.133	0.000
Was exposed to another science awareness program this year	0.202	0.204	0.200	0.769
Has been exposed to other science awareness programs in the past	0.324	0.299	0.349	0.007

Notes: The summary statistics are computed from the post-treatment student survey that was conducted in all treated and control classes between one and six months after the visits.

Table K13: Grade 10 – Spillovers Measured on Control Group Students, Using the Share of Classes Assigned to Treatments in School and Grade Level as an Instrument for Having Heard about the Program

	Grade 10 students (control group)					
	Girls			Boys		
	OLS (1)	2SLS (2)	2SLS (3)	OLS (4)	2SLS (5)	2SLS (6)
First stage (dependent variable: 'has heard about classroom interventions')						
Share of treated students in Grade level		0.600***	0.552**		0.195	0.098
		(0.192)	(0.231)		(0.125)	(0.151)
Outcomes (coefficient on 'has heard about classroom interventions')						
Negative perceptions of jobs in science (index)	0.017	0.113	0.107	−0.035	−1.189	−1.397
	(0.056)	(0.362)	(0.509)	(0.054)	(1.300)	(3.959)
Men are more gifted in math: True	0.041*	0.174	0.291	−0.016	0.211	−0.199
	(0.021)	(0.143)	(0.220)	(0.029)	(0.684)	(2.130)
Brains of women and men are different: True	0.022	0.021	−0.058	0.020	−0.247	−1.890
	(0.026)	(0.181)	(0.274)	(0.026)	(0.398)	(3.016)
Women like science less than men: True	0.007	−0.090	0.082	0.031	−0.010	0.117
	(0.022)	(0.128)	(0.177)	(0.027)	(0.385)	(1.113)
More men in science-related jobs: True	0.054**	−0.063	0.276	0.026	−0.180	0.431
	(0.025)	(0.174)	(0.268)	(0.030)	(0.549)	(1.862)
Women are discriminated in science: True	0.026	−0.020	0.172	0.070**	−0.910	1.704
	(0.036)	(0.226)	(0.379)	(0.031)	(0.975)	(2.916)
Taste for science subjects (index)	0.044	0.148	0.257	0.132***	1.505	7.186
	(0.054)	(0.423)	(0.656)	(0.045)	(1.414)	(11.467)
Low math self-concept (index)	0.098**	0.209	−0.282	0.011	−0.535	−5.306
	(0.047)	(0.338)	(0.516)	(0.041)	(1.648)	(9.757)
Interest in science-related careers (index)	0.146**	−0.266	−0.158	0.071	1.471	6.498
	(0.059)	(0.469)	(0.689)	(0.056)	(1.759)	(11.430)
Grade 11 : STEM track	−0.026	−0.039	−0.076	0.031	0.706	4.216
	(0.024)	(0.263)	(0.397)	(0.027)	(0.871)	(7.046)
Grade 11 : General science track	−0.024	−0.109	−0.123	0.008	−0.565	0.299
	(0.023)	(0.218)	(0.320)	(0.025)	(0.531)	(1.435)
Grade 11 : Technical STEM track	−0.002	0.069	0.046	0.023	1.270	3.916
	(0.008)	(0.133)	(0.206)	(0.018)	(1.006)	(6.554)
Controls						
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Number of Grade 10 students in school	No	No	Yes	No	No	Yes
N	3,805	3,805	3,805	3,324	3,324	3,324

Notes: Each value corresponds to a different model estimated on the population of Grade 10 students in the control group, based on students' responses to the post-treatment survey. Columns 1 and 4 report the coefficient of an OLS regression of each outcome on the dummy variable 'has heard about the program'. The other columns report the first stage and 2SLS estimates from a model in which the share of treated Grade 10 classes in the high school is used as an instrument for having heard about the program. All regressions control for students' characteristics. Column 3 and 6 further control for the total number of Grade 10 classes in the school. Standard errors (in parentheses) are adjusted for clustering at the classroom level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table K14: Grade 12 (Science Track) – Spillovers Measured on Control Group Students, Using the Share of Classes Assigned to Treatments in School and Grade Level as an Instrument for Having Heard about the Program

	Grade 12 (science track) students (control group)					
	Girls			**Boys**		
	OLS (1)	2SLS (2)	2SLS (3)	OLS (4)	2SLS (5)	2SLS (6)
First stage (dependent variable: 'has heard about classroom interventions')						
Share of treated students in Grade level		0.542**	0.167		0.206	0.082
		(0.237)	(0.317)		(0.201)	(0.247)
Outcomes (coefficient on 'has heard about classroom interventions')						
Negative perceptions of jobs in science (index)	−0.093	0.378	1.178	−0.133**	−1.564	−3.144
	(0.068)	(0.486)	(3.329)	(0.067)	(1.546)	(11.163)
Men are more gifted in math: True	0.017	0.203	0.475	0.020	−0.232	1.013
	(0.023)	(0.180)	(1.423)	(0.031)	(0.826)	(3.356)
Brains of women and men are different: True	0.032	0.379*	1.239	0.006	0.738	1.563
	(0.021)	(0.222)	(2.292)	(0.021)	(0.800)	(5.019)
Women like science less than men: True	−0.002	−0.018	−0.770	−0.018	0.542	5.597
	(0.019)	(0.097)	(1.433)	(0.024)	(0.789)	(25.024)
More men in science-related jobs: True	0.033	0.147	0.449	−0.003	−0.304	1.032
	(0.029)	(0.255)	(1.805)	(0.028)	(0.655)	(4.119)
Women are discriminated in science: True	0.042	−0.207	−0.433	0.036	−0.366	−3.276
	(0.033)	(0.308)	(1.765)	(0.030)	(0.762)	(9.831)
Taste for science subjects (index)	0.096	0.584	4.246	0.140**	−0.734	0.300
	(0.066)	(0.449)	(6.818)	(0.069)	(1.806)	(6.460)
Low math self-concept (index)	−0.005	0.007	2.085	−0.039	−1.853	−7.813
	(0.050)	(0.509)	(5.041)	(0.052)	(2.165)	(23.447)
Interest in science-related careers (index)	0.020	0.143	−1.733	0.091	−0.003	−4.657
	(0.060)	(0.428)	(3.672)	(0.062)	(1.423)	(14.983)
Undergraduate major in STEM	−0.035	0.173	0.964	−0.040	−0.333	−4.360
	(0.030)	(0.242)	(1.961)	(0.032)	(0.755)	(13.018)
Selective STEM undergraduate major	−0.001	0.132	1.005	0.010	0.031	−1.160
	(0.020)	(0.182)	(1.881)	(0.028)	(0.617)	(4.168)
Male-dominated STEM undergraduate major	−0.008	0.192	1.126	−0.032	0.633	−3.331
	(0.023)	(0.223)	(2.359)	(0.031)	(1.016)	(10.007)
Controls						
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Number of Grade 12 (science track) students in school	No	No	Yes	No	No	Yes
N	1,453	1,453	1,453	1,466	1,466	1,466

Notes: Each value corresponds to a different model estimated on the population of Grade 12 students in the control group, based on students' responses to the post-treatment survey. Columns 1 and 4 report the coefficient of an OLS regression of each outcome on the dummy variable 'has heard about the program'. The other columns report the first stage and 2SLS estimates from a model in which the share of treated Grade 10 classes in the high school is used as an instrument for having heard about the program. All regressions control for students' characteristics. Column 3 and 6 further control for the total number of Grade 12 classes in the school. Standard errors (in parentheses) are adjusted for clustering at the classroom level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table K15: Grade 10 – Spillovers Measured on Control Group Students, Proxied by the Share of Classes Assigned to Treatment in School and Grade Level

	Grade 10 students (control group)			
	Girls		**Boys**	
	(1)	(2)	(3)	(4)
Negative perceptions of jobs in science (index)	0.201	0.319	−0.157	−0.043
	(0.211)	(0.273)	(0.173)	(0.255)
Men are more gifted in math: True	0.114	0.173*	0.032	−0.012
	(0.075)	(0.094)	(0.122)	(0.167)
Brains of women and men are different: True	0.026	−0.014	−0.027	−0.141
	(0.104)	(0.132)	(0.076)	(0.099)
Women like science less than men: True	−0.082	−0.048	−0.047	−0.068
	(0.071)	(0.099)	(0.075)	(0.116)
More men in science-related jobs: True	−0.076	0.038	−0.071	−0.029
	(0.099)	(0.144)	(0.101)	(0.167)
Women are discriminated in science: True	−0.031	0.055	−0.144	0.139
	(0.126)	(0.163)	(0.108)	(0.166)
Taste for science subjects (index)	0.182	0.302	0.315*	0.737***
	(0.238)	(0.320)	(0.186)	(0.268)
Low math self-concept (index)	0.210	0.022	0.027	−0.315
	(0.211)	(0.287)	(0.302)	(0.424)
Interest in science-related careers (index)	−0.150	−0.051	0.229	0.467
	(0.257)	(0.342)	(0.222)	(0.300)
Grade 11 : STEM track	0.123	0.164	0.154	0.388**
	(0.137)	(0.177)	(0.127)	(0.161)
Grade 11 : General science track	0.032	0.067	−0.091	0.026
	(0.109)	(0.135)	(0.100)	(0.125)
Grade 11 : Technical STEM track	0.091	0.096	0.245*	0.362**
	(0.066)	(0.087)	(0.128)	(0.160)
Controls				
Student characteristics	Yes	Yes	Yes	Yes
Number of Grade 10 students in school	No	Yes	No	Yes
N	3,805	3,805	3,324	3,324

Notes: Each value corresponds to a different model estimated on the population of Grade 10 students in the control group, based on students' responses to the post-treatment survey. The columns report the coefficient of an OLS regression of each outcome on the share of treated Grade 10 classes in the school, which is used as a proxy for potential spillovers. All regressions control for students' characteristics. Columns 2 and 4 further control for the total number of Grade 10 classes in the school. Standard errors (in parentheses) are adjusted for clustering at the classroom level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table K16: Grade 12 (Science Track) – Spillovers Measured on Control Group Students, Proxied by the Share of Treated Classes by School and Grade Level

	Grade 12 (science track) students (control group)			
	Girls		Boys	
	(1)	(2)	(3)	(4)
Negative perceptions of jobs in science (index)	0.184	0.137	−0.285	−0.126
	(0.213)	(0.301)	(0.196)	(0.211)
Men are more gifted in math: True	0.152*	0.181	0.004	0.196
	(0.079)	(0.115)	(0.145)	(0.158)
Brains of women and men are different: True	0.210***	0.219**	0.159	0.167
	(0.076)	(0.105)	(0.102)	(0.147)
Women like science less than men: True	0.001	−0.093	0.078	0.223
	(0.048)	(0.060)	(0.105)	(0.192)
More men in science-related jobs: True	0.008	−0.077	−0.133	−0.123
	(0.136)	(0.226)	(0.118)	(0.196)
Women are discriminated in science: True	−0.187	−0.239	−0.166	−0.405**
	(0.147)	(0.222)	(0.134)	(0.184)
Taste for science subjects (index)	0.192	0.316	−0.133	0.014
	(0.259)	(0.359)	(0.283)	(0.406)
Low math self-concept (index)	0.004	0.317	−0.302	−0.305
	(0.259)	(0.402)	(0.228)	(0.356)
Interest in science-related careers (index)	0.130	−0.059	−0.033	−0.373
	(0.214)	(0.314)	(0.269)	(0.340)
Undergraduate major in STEM	0.105	0.091	0.014	−0.146
	(0.100)	(0.150)	(0.131)	(0.211)
Selective STEM undergraduate major	0.103	0.168*	0.013	−0.120
	(0.069)	(0.101)	(0.107)	(0.153)
Male-dominated STEM undergraduate major	0.106	0.136	0.156	−0.128
	(0.077)	(0.115)	(0.123)	(0.168)
Controls				
Student characteristics	Yes	Yes	Yes	Yes
Number of Grade 12 (science track) students in school	No	Yes	No	Yes
N	1,453	1,453	1,466	1,466

Notes: Each value corresponds to a different model estimated on the population of Grade 12 (science track) students in the control group, based on students' responses to the post-treatment survey. The columns report the coefficient of an OLS regression of each outcome on the share of treated Grade 10 classes in the school, which is used as a proxy for potential spillovers. All regressions control for students' characteristics. Columns 2 and 4 further control for the total number of Grade 12 (science track) classes in the school. Standard errors (in parentheses) are adjusted for clustering at the classroom level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix L Effects on Academic Performance in Grade 12

Table L17: Treatment Effect on *Baccalauréat* Performance, Grade 12 Students

	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)
Baccalauréat percentile rank in math	46.20	0.693	0.469	47.24	1.661	0.105
		(0.957)	[0.611]		(1.024)	[0.175]
Baccalauréat percentile rank in French	54.35	−0.051	0.964	43.30	−0.331	0.680
		(1.113)	[0.964]		(0.803)	[0.680]
Baccalauréat percentile rank	53.21	−1.121	0.293	47.05	1.712*	0.100
		(1.066)	[0.611]		(1.040)	[0.175]
Obtained the *Baccalauréat*	0.928	−0.010	0.334	0.875	−0.005	0.623
		(0.010)	[0.611]		(0.010)	[0.680]
N		2,827			2,924	

Notes: This table reports treatment effects estimates on Grade 12 students' *Baccalauréat* grades, separately by gender. Each row corresponds to a different model, based on students' responses to the post-treatment survey. Columns 1 and 4 report the average value for students in the control group. Columns 2 and 5 report the coefficient on a "facilitator visit" dummy in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report both the standard p-value and, in brackets, the p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix M Heterogenous Effects on Educational Choices

M.1 By Student Performance in Math

Table M18: Enrollment Status the Following Year, by Level of Performance in Math

	Girls			**Boys**		
	Treatment effect (LATE) by level of performance in math			Treatment effect (LATE) by level of performance in math		
	Below median	Above median	Difference: p-value [q-value]	Below median	Above median	Difference: p-value [q-value]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Grade 10						
Grade 11: Science Track	−0.020	0.004	0.308	−0.019	0.009	0.330
	(0.015)	(0.018)	[0.462]	(0.021)	(0.018)	[0.648]
Grade 11: General STEM Track	−0.005	0.002	0.759	−0.004	0.016	0.431
	(0.014)	(0.019)	[0.759]	(0.016)	(0.018)	[0.648]
Grade 11: Technical STEM Track	−0.015**	0.002	0.062	−0.015	−0.007	0.702
	(0.007)	(0.004)	[0.186]	(0.016)	(0.011)	[0.702]
N	3,534	3,707		3,148	3,311	
Panel B. Grade 12 (science track)						
College: Undergraduate major in STEM	0.010	0.032	0.537	−0.037	0.031	0.089
	(0.020)	(0.025)	[0.537]	(0.026)	(0.029)	[0.192]
College: Selective STEM track	0.001	0.066***	0.015	−0.012	0.040	0.127
	(0.013)	(0.020)	[0.044]	(0.018)	(0.027)	[0.192]
College: Major in male-dominated STEM	0.025	0.046**	0.499	−0.002	0.024	0.498
	(0.018)	(0.022)	[0.537]	(0.025)	(0.028)	[0.498]
N	1,544	1,482		1,497	1,544	

Notes: This table reports treatment effects estimates on students' educational outcomes in the academic year following the classroom interventions, i.e., 2016-2017, separately by grade level, gender, and by students' academic performance in math. Each row corresponds to a different model, based on administrative data on student enrollment outcomes. Student academic performance in math is measured from the grades they obtained on the final math exam of the *Diplôme national du Brevet* at the end of middle school for Grade 10 students, and from the grades obtained on the *Baccalauréat* math final exam for Grade 12 students. Columns 1 and 2 (for girls) and 4 and 5 (for boys) report the coefficients on the interactions between between a "facilitator visit" and dummies for the student being either below (columns 1 and 4) or above (columns 2 and 5) the median level of performance in math, in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). The regression controls for school fixed effects to account for the fact that randomization was stratified by school. Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report the standard p-value for the difference in the treatment effects estimates for students above vs. below the median performance in math. The p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes are reported in brackets, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

M.2 By Facilitator Background

Table M19: Enrollment Status the Following Year, by Facilitator Background

	Girls			Boys		
	Treatment effect (LATE) by facilitator background			Treatment effect (LATE) by facilitator background		
	Researcher	Professional	Difference: p-value [q-value]	Researcher	Professional	Difference: p-value [q-value]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Grade 10						
Grade 11: Science Track	0.000	−0.001	0.975	−0.025	0.009	0.218
	(0.025)	(0.026)	[0.975]	(0.026)	(0.024)	[0.589]
Grade 11:General STEM Track	0.008	0.005	0.928	−0.012	0.005	0.589
	(0.025)	(0.026)	[0.975]	(0.029)	(0.024)	[0.589]
Grade 11: Technical STEM Track	−0.008	−0.006	0.787	−0.014	0.004	0.470
	(0.006)	(0.006)	[0.975]	(0.020)	(0.022)	[0.589]
N	3,237	4,004		2,890	3,569	
Panel B. Grade 12 (science track)						
College: Undergraduate major in STEM	0.004	0.032	0.298	−0.016	0.022	0.344
	(0.026)	(0.022)	[0.447]	(0.037)	(0.031)	[0.425]
College: Selective STEM track	0.001	0.052**	0.048	0.002	0.033	0.424
	(0.022)	(0.022)	[0.146]	(0.033)	(0.033)	[0.425]
College: Major in male dominated STEM	0.024	0.042*	0.546	−0.023	0.047	0.087
	(0.025)	(0.022)	[0.546]	(0.037)	(0.032)	[0.261]
N	1,180	1,647		1,312	1,612	

Notes: This table reports treatment effects estimates on students' educational outcomes in the academic year following the classroom interventions, i.e., 2016-2017, separately by grade level, gender, and by background of female facilitator who visited the classroom (professional or researcher). Each row corresponds to a different model, based on administrative data on student enrollment outcomes. Columns 1 and 2 (for girls) and 4 and 5 (for boys) report the coefficients on the interactions between between a "facilitator visit" and dummies for the facilitator being either a researcher (columns 1 and 4) or a professional (columns 2 and 5), in a regression of each outcome on treatment, using random assignment as an instrument for receipt of treatment and interpreting the resulting estimand as a Local Average Treatment Effect (LATE). Standard errors (shown in parentheses) are adjusted for clustering at the unit of randomization (classroom). Columns 3 and 6 report the standard p-value for the difference in the treatment effects estimates for students above vs. below the median performance in math. The p-value (q-value) adjusted for multiple hypotheses testing across variables belonging to the same "family" of outcomes are reported in brackets, 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.