CENTER FOR RESEARCH ON PENSIONS AND WELFARE POLICIES

**Working Paper 165/16**

# TEACHER CHARACTERISTICS, STUDENT BELIEFS AND THE GENDER GAP IN STEM FIELDS

**Dario Sansone**

October 2016

# Teacher Characteristics, Student Beliefs and the Gender Gap in STEM Fields[♠]

Dario Sansone[1]

October 2016

**Abstract**

This paper investigates how to reduce the gender gap in STEM fields in the US. It quantifies the impact of high-school teachers' gender, beliefs and behavior on students' beliefs about girls' abilities in math and science. Furthermore, it shows that such beliefs affect female students' decision to take advanced math and science classes in high school, as well as their intentions to choose a STEM major once freshman in college.

**Keywords**: gender gap, high school, STEM, women in science, beliefs.

**JEL**: J16, I20, D83.

---

♠ I am grateful to Francis Vella, Garance Genicot, Laurent Bouton, Martin Ravallion, Debraj Ray, Arthur van Soest, Jim Albrecht, Susan Vroman, Joseph Altonji, Mariacristina Rossi, Luca Flabbi, Arik Levinson, Mary Ann Bronson, Franco Peracchi, Miikka Rokkanen, Jeff Strohl, Rasmus Lentz, Austin Nichols, Federico Bugni, Billy Jack, Anil Nathan, Allison Stashko, Pooya Almasi, Michael Packard, Givi Melkadze and participants to the Gender and STEM workshop (Collegio Carlo Alberto), the IAFFE conference (NUI Galway), the Gender Economics and Society Summer School (ITILO), and the PhD seminar (Georgetown University) for their helpful comments. I am also grateful to John Rust and Judith House for their technical support. The usual caveats apply.

1 Georgetown University (Department of Economics, ICC 580, 37th and O Streets, NW, Washington DC 20057-1036, USA. Phone: +12026875601). E-mail: ds1289@georgetown.edu

## 1. Background

This paper studies how student-teacher interactions in high school shape female students' decisions to take advanced math and science classes in high school and to major in a STEM (Science, Technology, Engineering and Mathematics) field once in college. In order to contextualize such analysis, this introduction provides background information on the gender gap in science and math for the US and it motivates the importance of increasing workers, particularly female ones, in STEM fields.

Historically, male students have outperformed female ones in math test scores. Nevertheless, such gender gap has been narrowing down in the National Assessment of Educational Progress (NAEP) mathematics scores since 1973[2] (National Center for Education, 2013). Similarly, the gender gap in the Programme for International Student Assessment (PISA) of 2012 for math and science was not statistically significant ((OECD, 2014), (OECD, 2015)). Furthermore, (Hyde, Lindberg, Linn, Ellis, & Williams, 2008) found a rather small gender gap using data available following the No Child Left Behind legislation.

Looking at course choice in high school, male students are more likely to take AP courses in mathematics, physics and computer science, while the opposite is true for biology and environmental science (National Science Board, 2014). These trends continue in tertiary education (National Science Foundation, 2015). Indeed, the proportion of women in science is quite high in psychology, bioscience and social science (except economics). However, women are still a minority in engineering, computer science, physics, economics, mathematics and statistics[3]. Furthermore, in some of these fields, like computer science, the proportion of women has decreased over time, while in the other fields there has not been a consistent increase in the proportion of degrees awarded to women. The figures are similar – or worse – at the graduate level. Focusing instead on attrition rates, (Chen, 2013) documented that 28% of undergraduate students chose a STEM major between 2003 and 2009, but 48% of them left the field by 2009. Among these, around half switched to a non-STEM major, while the others dropped out of college: relatively more women than men switched major, while more men than women dropped out of college. As a result, researchers may wonder whether these high attrition rates are because male and female students do not have a good preparation in these subjects from high school.

These trends translate in occupational differences: 50% of scientists and engineers are white men. In academia, women represented one-fourth of full-time, full professors in science and engineering in 2013 (12% in economics (Economist, 2015)). This underrepresentation is clear also in the Silicon Valley ((Google, 2015), (Apple, 2015), (Williams, 2014)).

---

[2] This gender gap among 17-years-old students has become smaller thanks to female students' gains, while male students had stable performances.

[3] Outside STEM, it is worth mentioning that the proportion of women with a business degree increased substantially between the '50s and the '80s and it is now slightly below 50% (NCES, 2015).

All these figures have generated a national debate about women in science ((Varma, 2010), (Eileen Pollack, 2013), (Blow, 2015), (Eillen Pollack, 2015)) and the shortage of STEM workers ((Sabot & Wakeman-Linn, 1991), (Rosen, 2013), (Leef, 2014)). The latter issue was raised in (Carnevale, Smith, & Melton, 2011) and (Executive Office of the President, 2012). The former argued that more and more sectors outside STEM were requiring individuals with STEM competencies, thus creating such a shortage. Along the same lines, the importance of women in STEM fields have been highlighted by several other scholars ((Page, 2008), (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010), (Corbett & Hill, 2015), (Sikdar, 2015))[4].

Finally, there is growing evidence that gender differences in math and science are not caused by genetic factors ((Guiso, Monte, Sapienza, & Zingales, 2008), (Fryer Jr & Levitt, 2010), (OECD, 2015), (Wheeling, 2015), (Friedman-Sokuler & Justman, 2016)). Although it is true that men and women approach complex mathematical problems in different ways, this does not imply that one gender has an advantage in learning advanced mathematics (Spelke, 2005)[5].

It is therefore possible to conclude that, although the gender gap in test scores has been declining and there is empirical evidence that women are not innately inferior to men in math and science, women tend to choose different courses in high school and to major less frequently in STEM fields, thus decreasing productivity in those sectors and aggravating the lack of STEM workers (as well as perpetuating the wage gender gap). This paper investigates whether it is possible to improve college readiness in math and science among female students and to increase the number of female students in STEM majors by changing their beliefs about girls' abilities in math and science through their high-school teachers.

[4] One may argue that a diverse group could be less productive because of coordination and communication problems. But the point here is different: a diverse group has the potentiality to be more productive than a homogeneous group. Which organization structure allows the full use of such potential is a different question and it is not addressed here.

[5] It is also interesting to note that (Leslie, Cimpian, Meyer, & Freeland, 2015) showed that gender imbalances are predominant in subjects in which practitioners believe that raw, innate talent is a key element for success in the field.

## 2. Conceptual Framework

This paper employs data in the US from the High School Longitudinal Study of 2009 (HSLS:09) to investigate the determinants of the gender gap in STEM fields. As also summarized by (Ceci, Ginther, Kahn, & Williams, 2014), this topic has been extensively analyzed in recent years by, among the others, (Guiso et al., 2008), (Mechtenberg, 2009), (Fryer Jr & Levitt, 2010), (Schneeweis & Zweimüller, 2012).

The specific channel that is investigated here is the impact of teachers' gender, beliefs and behaviors on students' beliefs about boys' and girls' abilities in math and science. A unique feature of this dataset is exploited: teachers, students and parents are asked to compare boys and girls in math and science. The aim is to explore if and how teachers affect students' responses. If the estimates showed that teachers' gender does drive the results, i.e. *ceteris paribus* female teachers lead fewer students to believe that boys are better than girls in math or science, the policy implications would be much different than if the estimates indicated that the key drivers are how the teachers behave in class, or the teachers' own ex-ante beliefs about boys' and girls' math and science skills.

This section of the paper has been motivated by the rapidly growing literature on the effect of the teacher gender on student performances. Indeed, (Dee, 2005) showed that teachers' race and gender have large impacts on their perceptions of students' performances and behavior. Similarly, (Carrell, Page, & West, 2010) found positive effects of female professors on high-performing female undergraduate students. Following these papers, several analyses have been conducted in primary and secondary schools ((Dee, 2007), (Holmlund & Sund, 2008), (Winters, Haight, Swaim, & Pickering, 2013), (Paredes, 2014), (Antecol, Ozkan, & Serkan, 2015), (Muralidharan & Sheth, 2016)), as well as universities ((Bettinger & Long, 2005), (Hoffmann & Oreopoulos, 2009), (Price, 2010), (Bottia, Stearns, Mickelson, Moller, & Valentino, 2015)). Although most of these studies have found positive effects of female teachers on female students' achievements, the overall mixed results indicate that the issue may be more complex. Indeed, female teachers represent a highly heterogeneous group, so it is unsurprising that the empirical conclusions have not been clear-cut. This paper contributes to this literature by differentiating between several teachers' characteristics. In fact, similarly to (Gunderson, Ramirez, Levine, & Beilock, 2012) and (Kramer et al., 2016), the detailed dataset allowed me to understand whether female students are more influenced by female teachers because of role model effect (or lower stereotype threat), or whether what really matters is not the teachers' gender, but their beliefs, how they treat the students and manage the classroom.

Using a between school-subject fixed effect, i.e. by comparing how math and science teachers affect students' beliefs, I show that the teacher gender is indeed pivotal. If a student (female or male) has a female teacher in 9$^{th}$ grade, she is less likely to believe that boys are better than girls in math or science. In addition to this, a similar effect on female students' beliefs is found when teachers listen and value students' ideas. On the other hand, I cannot reject the null hypothesis that teachers' beliefs have a null impact on students' beliefs.

The second part of the paper looks at the effect of these female students' beliefs on their decision to take advanced math and science classes in high school, as well as their intentions to choose a STEM major once in college. If the estimates proved that female students who do not believe that boys are better than girls in math or science were more likely to take more math and science classes in high school and major in a STEM field once in college, then policy-makers may increase the number of female STEM undergraduates, as well as improving their college readiness, by increasing their confidence in girls' math abilities through their teachers in high school.

The idea behind this model is that students choose their majors by comparing expected costs and benefits of each field. Similarly to gender roles (Vella, 1994), beliefs enter into this decision mechanism by affecting expectations. For instance, if a female student believed that boys were better than girls in science, she may expect a more hostile environment because of fewer women in STEM fields, thus increasing the expected cost of choosing such major. Similarly, such belief may also lower the student's confidence in her scientific abilities, thus also increasing expected costs.

I am not claiming that this is the only channel through which the gap in STEM fields may be filled. The goal here is to offer an explanation which is a complement rather than a substitute to those already stressed in the literature. In other words, as (Ceci et al., 2014), I argue that pre-college factors may drive the gender gap in math-intensive fields. Beliefs are particularly important in high school: whether a female student in 9$^{th}$ grade plans to enroll in advance math or science classes may depend not only on her expected earnings in or out of science after 10 years, but also on how she compares boys and girls in math and science, that is on her confidence in women's abilities and on her expectation about the learning environment. However, I do not claim that this is the only way to tackle this issue. High school teachers may play an important role in shaping the career choices of their students, but this does not rule out that female students take into account other variables when deciding their future path. Indeed, as summarized in (Altonji, Arcidiacono, & Maurel, 2015), several authors have looked at the characteristics of the labor markets, as well as individual preferences, in order to explain these differences in major and occupational choices ((Turner & Bowen, 1999), (Weinberger & Leggon, 2004), (Zafar, 2013), (Gemici & Wiswall, 2014), (Goldin, 2014), (Reuben, Sapienza, & Zingales, 2014), (Bronson, 2015) .

The general framework is in the vein of (Coate & Loury, 1993): negative beliefs may constitute a self-fulfilling prophecy. In other words, lack of role models, adverse class and work environments as well as different expectations for girls and boys may lead more female students to believe that it is indeed true that boys are better than girls in math and science. This may affect their choices about coursework in high school and major in college, thus causing fewer women to select a STEM major, and those who do choose that path would typically be not ready since they took fewer advanced math and science classes than their male classmates in high school. The final result would be a shortage of highly-qualified women in STEM, which would confirm the lower expectations held for women and lead to even worse conditions in term of role models and environments. The aim of this paper is to prove that such a vicious cycle between beliefs and performances exists, and that it is possible to break this cycle by improving teacher-student interactions[6].

Using again a between-subject fixed effect, I show that if a 9th grade female student believes that boys are better than girls in math or science, she is less willing to take advance classes in those subject while in high school. Furthermore, if a female student believes that boys are better than girls in science, she is less likely to declare a major in a STEM field once in college. I use a linear probability model which tackle any omitted variable issues by including several socio-demographic controls, as well as family information, school characteristics, and student choices and performances in high school. This result is also confirmed by a linear IV model and additional robustness checks.

[6] Put differently, the underlining theoretical model has multiple equilibria where the current "bad" equilibrium has female underrepresentation in STEM fields. If this equilibrium were stable, a marginal improvement as the one discussed in this paper may not be enough. However, in case of instability, even a small change may initiate a virtuous circle which would eventually lead to a more gender-balanced equilibrium. Similar cases, such as the gradual increasing participation of women to tertiary education and the labor force, suggest that these initial equilibria are indeed unstable (probably due to their inherited inefficiency and unfairness).

## 3. Data

The High School Longitudinal Study of 2009 (HSLS:09) is a panel micro database including around 26,000 students in 9th grade from about 940 participating schools in 2009. The survey design has two levels: first, schools were selected at the national level (both private and public). Second, around 30 students in each school were randomly selected among 9th grades[7]. Among eligible students, around 21,440 students responded.

In the first round, information was collected from the selected 9th graders, their parents, math and science teachers, school administrators and lead school counselors. The parent questionnaire was completed by the parent or guardian most familiar with the 9th grader's school situation and experience. If the 9th grader had more than one science or math teacher, one teacher per subject was randomly selected among those provided. The students were interviewed between September 2009 and April 2010. The first follow-up was in the spring of 2012, while a brief update was conducted in 2013 (summer and fall) to record students' postsecondary plans. In 2012 students, parent, school administrators and counselors were interviewed again, but this wave did not include new questionnaires for teachers. Finally, in 2013 only students and parents were interviewed. The dataset with the 2013 Follow-up became publicly available in June 2015.

A math assessment was administered to the students in 9th grade (2009) and in 11th grade (2012). Data are also available from the student transcripts including their GPA, their AP class grades, their SAT, and the number of credits taken in each subject during high school.

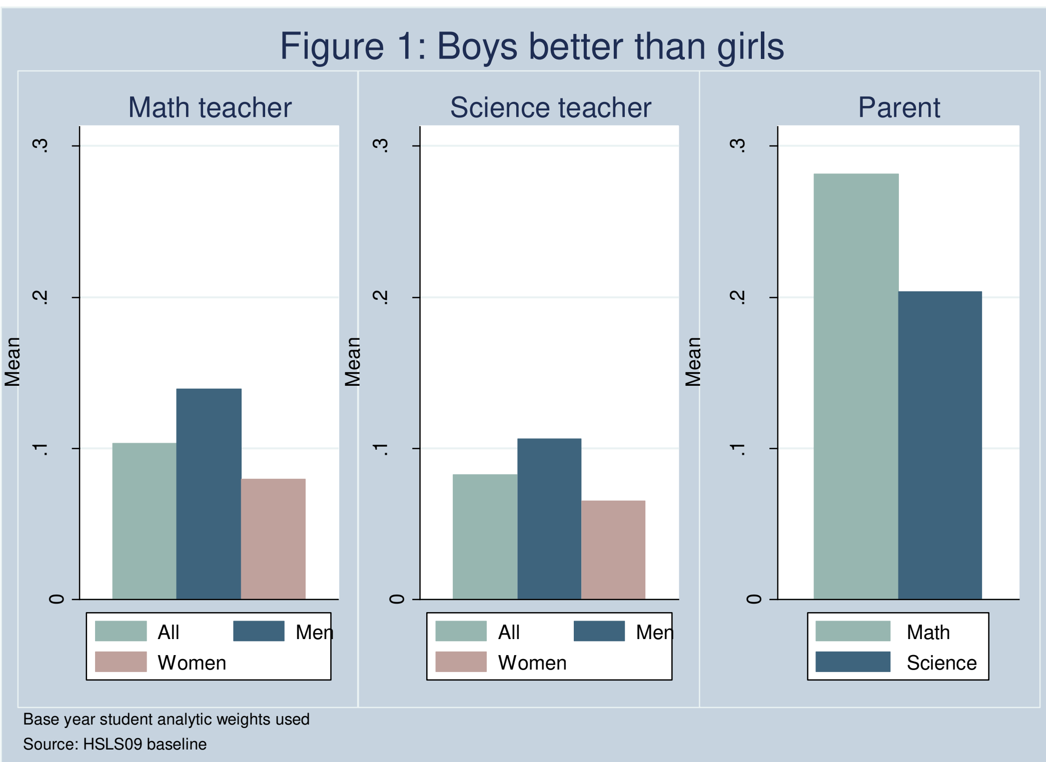
Additional documentation about the HSLS:09 can be found in the online training modules (NCES, 2016) and in (Ingels et al., 2011), (Ingels et al., 2014) and (Ingels et al., 2015). The HSLS:09 have been used in a number of studies loosely related to this paper by (Degner, 2013) (Jackson, 2013), and (Wagstaff, 2014).

---

[7] The complex survey design has been taken into account in the empirical analysis by clustering the standard errors at the school level. Such analysis has been conducted using Stata 14. The dataset also includes the student analytic weights for the base year survey and the longitudinal study. Following (Solon, Haider, & Wooldridge, 2015), such weights have been used for the descriptive statistics. However, in this paper the sampling is independent of the dependent variables conditional on the explanatory variables, so using weights to correct for endogenous sampling does not seem appropriate here. Moreover, using weights in order to estimates average partial effects in case of heterogeneous effects is usually insufficient. Therefore, rather than weighting, when I suspected heterogeneity in this paper I tried to analyze it by adding interaction terms or by focusing the estimation on a sub-sample. Finally, the general view is that is more conservative to report heteroscedasticity-robust standard errors rather than using weights to obtain more precise estimates under heteroscedasticity. To conclude, I do not find sufficient reasons to justify weighting in this paper when I estimate causal effects.

## 4. Descriptive Statistics

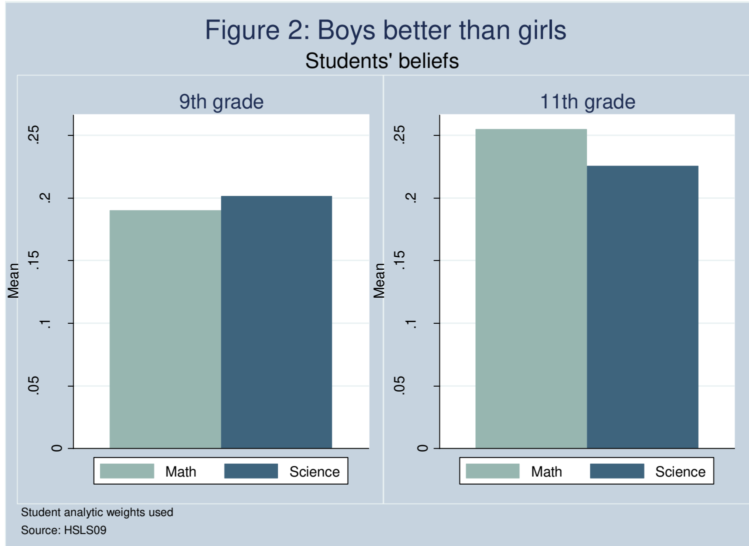
The key variables in this analysis are the respondents' beliefs about boys' and girls' abilities in math and science[8]. As far as high school teachers are concerned (Figure 1), around 10.6% of the math teachers believed that boys are better than girls in math. Furthermore, this opinion was more common among male math teachers (14%). Nevertheless, it is interesting to note that there were also some female math teachers (8%) who believed that boys are better in math. In science, around 8.3% of the science teachers believed that boys are better than girls in science. Male science teachers were slightly more likely to hold such belief (10.6%), although this was also true for some female teachers (6.5%)[9]. It is also important to highlight that almost 60% of math teachers and 57.5% of science teachers in the sample were female. On the other hand, around 28.2% of the parents believed that boys are better than girls in math, while 20.4% supported this idea in science.




[8] The Appendix contained a detailed description of the variables used in this and the following empirical sections, as well as all additional robustness checks. When not reported, tables are available upon request. In addition to this, it is important to note that all sample size numbers have been rounded to the nearest 10 for security reasons.

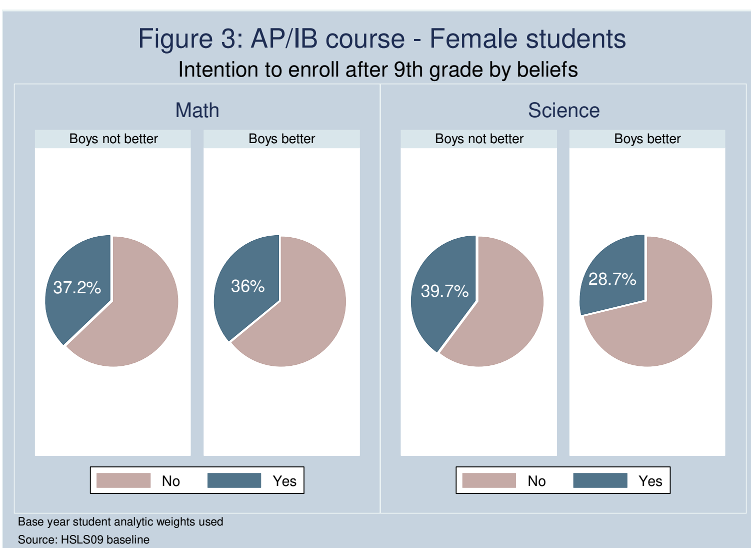
[9] These gender differences across teachers are statistically significant.

Looking instead at the student data (Figure 2), when students were in 9th grade, around 19% of the students believed that boys are better than girls in math, while 20.2% supported this idea in science[10]. Things got worse in 11th grade: around 25.5% of the students believed that boys are better than girls in math, while 22.6% had the same opinion in science.



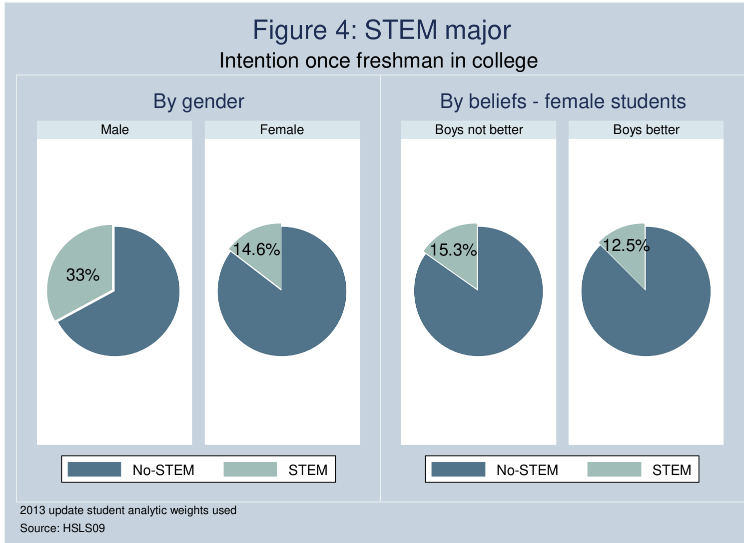

It is also important to note that beliefs tend to be persistent: more than 80% (85% among girls) of the students who did not believe that boys are better than girls in science in 9th grade had the same opinion in 11th grade (while 62% of those who believed that boys are better changed their opinion). This fact offers an additional incentive to find a way to change these beliefs early on: given that relatively few students switched between 9th and 11th grade from thinking that boys are not better than girls in math or science to the opposite, it seems worth trying to convince students at the beginning of high school that indeed boys are not better than girls in these subjects. This may also push female students to take advance classes in math and science during the first years of high school.

In fact, looking at how many students planned to take advanced math and science classes in high school, around 25% of students intended to enroll in an AP course. This figure was similar in math and science and for boys and girls. However, it seems that more female students were unsure about their future plans. In addition to this, it is worth mentioning that girls who believed that boys are better than girls in math or science were less likely to enroll in advanced classes (Figure 3). This was true especially in science: almost 40% of girls who did not believe that boys were better than girls in science intended to enroll in an AP or IB science class, while this was true only for 28.7% of the female students who had the opposite belief.




[10] Looking at female students only, the percentage was almost 15% in math and 17% in science.

Finally, another outcome variable of interest is the proportion of students who decided to major in a STEM field (Figure 4). The key startling fact is the big gender gap: among the students in the sample who planned to continue in a STEM field after their freshman year in college, only one third was female. Nevertheless, the proportion of female students interested in pursuing a scientific education was higher among those who did not believe while they were in high school that boys are better than girls in science. Such difference is small in absolute terms (around 3 percentage points), but rather substantial in relative terms.




These stylized facts are simple correlations and could actually reflect spurious relationships due to omitted variables. Therefore, it is necessary to verify whether there is indeed a causal link between students' beliefs and their subsequent choices in high school and college, and whether high school teachers can change such beliefs. The rest of the paper uses several econometric techniques to test the existence of these causal relationships.

## 5. Empirical Section I: Student-Teacher Interaction

### 5.1 Econometric Framework

I employ an econometric technique pioneered by (Dee, 2005) and employed with a binary dependent variable by (Gershenson, Holt, & Papageorge, 2016). This identification strategy exploits the fact that for each interviewed student in $9^{th}$ grade it is possible to obtain information about her math (M) and science (N) teachers. The specification is run separately for male and female students. The starting point is the following regressions:

$$b_{is} = z'_{is}\beta_1 + \mu_i + \alpha_s + \varepsilon_{is}\,, \forall s \in \{M, N\}$$

The dependent variable is whether or not student $i$ thinks that boys are better than girls in subject $s$. This depends on the math/science teacher's characteristics for student $i$ ($z_{Mi}$ and $z_{Ni}$), observable and unobservable student fixed-effects ($\mu_i$), subject fixed-effect ($\alpha_s$), and the error terms ($\varepsilon_{Mi}$ and $\varepsilon_{Ni}$).

Note that the matrix Z contains the key regressors, i.e. teachers' gender, beliefs and behavior in class. One immediate issue is that some additional teacher characteristics may affect student's belief and be correlated with those regressors, thus leading to an omitted variable bias. In order to tackle this concern, several teacher indicators have been added as controls. Indeed, the regressions include whether the teacher has a graduate degree, if the teacher majored in a STEM fields, as well as the number of year (and its squared term) of teaching experience in the subject.

Given this panel data specification, it is possible to control for unobservable variables that are constant across subjects at the individual level by taking the difference between the two equations. This allows me to obtain first difference (FD) estimates, where the difference is not between time – as usually done in panel data – but between teachers. Moreover, it is worth stressing that there is no need to impose that the impact of the math teacher is the same of the science teacher: it is possible to add heterogeneity by interacting teacher characteristics with the subject fixed-effect. In other words, it is possible to obtain estimates of $\beta_{1s}$ rather than just $\beta_1$.

As in (Gershenson et al., 2016), a linear probability model is preferred even if the dependent variable is a binary outcome. In addition to the advantages highlighted in (Joshua D Angrist, 2001), (J. D. Angrist & Pischke, 2009), this choice has been made since in a linear model is straightforward to add fixed-effects and the coefficients can be interpreted as average partial effects. A simple logit or probit model would not allow the inclusion of $\mu_i$ with only two observations for student because of the incidental parameter problem. An alternative approach would have been to estimate a conditional logit model. However, since the distribution of the fixed effects is unknown, it would have not been possible to estimate the average partial effects in this case, but only the effect of the regressors on the log-odds ratio[11].

[11] See (Wooldridge, 2010). Estimation results for the conditional logit model are qualitatively similar to the linear ones.

It is worth stressing the fact that the student fixed-effect controls for all variables which affect student beliefs and are constant across subjects. This does not only include student individual characteristics such as race or cognitive/non-cognitive skills, but also school characteristics, for instance whether the school is private or public or its gender composition, and family background, i.e. income, parental education, household demographics and so on. Nevertheless, in order to be more confident in the results, I have added one variable which changes across individuals and subjects at the household level: how parents compare boys and girls in math and science. In other words, the following equation has been estimated, where $x_{is}$ includes parent beliefs.

$$b_{is} = z'_{is}\beta_1 + x'_{is}\beta_2 + \mu_i + \alpha_s + \varepsilon_{is}, \forall s \in \{M, N\}$$

One may also worry about teachers misreporting their opinion when asked to compare boys and girls in math and science. However, as long as such measurement error is the same for math and science teachers and it has the same impact on student beliefs, it would be differenced out in the final equation. In fact, there is no reason to believe that math teacher may misreport more or less often than science teacher[12]. Furthermore, note that such concern should not be applied to the indicators of teachers' behaviors in class since they have been derived from the student questionnaire.

Another concern is that part of the unobservable student characteristics ($\mu_i$) may actually be subject specific (i.e. $\mu_i+\delta_{Mi}$ and $\mu_i+\delta_{Si}$), thus the FD equation would not control for such subject-specific student unobservables. First of all, it is possible to argue that, while these subject specific components may be important when comparing hard science with humanities, it does not seem that between math and science there is a substantial difference. For instance, as reported in (Patterson & Kobrin, 2012), there is a high correlation between the SAT scores in Math and Chemistry (0.756) or Physics (0.755). In addition to this, it is possible to control for students' course-s grade in the previous academic year, i.e. 8th grade[13].

It is also worth mentioning that around one fourth of students in the sample provided different answers about boys' and girls' abilities in math and science, thus there is enough variability to obtain estimates using variation within individuals. Moreover, the fact that there is such a variability provides additional evidence supporting the empirical strategy of splitting math and science beliefs. In other words, people provide different answers when they are asked to compare boys and girls in math or science, so it would not be appropriate to combine such beliefs.

[12] Likewise, the non-response rates are similar across subjects: around 15% of the students did not respond in 9th grade when they were asked to compare boys and girls in math and science. The non-response rates are also similar for parents in both subjects (around 32%). It is also important to note that at the individual level most of the students whose beliefs are missing in math also have a missing for science. The same can be said for the parents. For teachers, there is a small difference: 29% of math teachers did not reply when asked to compare boys and girls in math, while 35% of science teachers did not reply when asked to compare boys and girls in science.

[13] However, notice that these controls are potentially endogenous since they depend on 8th grade teachers' characteristics, which may affect students' beliefs in 9th grade as well. Therefore, such specification is not the main equation of interest, but it is added as a robustness check.

### 5.2 Empirical Results

The results are shown in Table 1 for female students and Table 2 for male students. Column 1 includes the basic regressors, additional teachers' characteristics are added in Column 2, while parental beliefs are taken into account in Column 3. The key result, which holds across all specifications, is that the gender of the teacher is pivotal in shaping students' beliefs. Indeed, looking at column 3, on average having a female teacher reduces the probability of believing that boys are better than girls in math/science by almost 6 percentage points for female students. A similar result holds for boys.

From a policy perspective, it is interesting to note that female teachers can also have an effect on boys. Indeed, these students will become husbands, fathers and colleagues. Therefore, changing their beliefs may have inter-generational effects as well as improving the workplace environment which is often reported as an obstacle to women participation in STEM fields.

As expected, parents also have an important impact on what their daughters and sons think. In addition to this, and consistent with the previous literature (Kramer et al., 2016), girls tend to be more sensitive to the learning environment. In fact, female students tend to be more confident about girls' abilities when their teachers create a positive classroom environment by listening to students' ideas[14].

On the other hand, teachers' beliefs do not seem to have a significant effect. This may be due to the lower variability in the regressor across subject. In fact, around 46% of the students in the sample have math and science teachers with different gender. Moreover, 22% of the students thinks that their math teacher listens to students' ideas but that this is not true for their science teacher (or vice versa). On the other hand, only 17.8% of the students have a math teacher who believes that boys are better than girls in math and a science teacher who does not believe that boys are better than girls in science (or vice versa).

There is also no evidence of heterogeneity between subjects: it seems that female teachers have a similar effect in math and science (Column 4). Moreover, if the interaction between teacher's gender and whether the teacher listens to student ideas is added in the specification for female students[15], its coefficient is positive and significant, thus indicating a certain degree of

[14] One concern about this variable is that it is reported by the student instead of the teacher. Therefore, there may be a latent variable which drives students' responses both when they are asked to compare boys and girls in math and science, as well as when they are asked whether their teacher listen to students' ideas. Omitting this regressor from the specification does not change the impact of the gender of the teacher. In addition to this, students are also asked in 9$^{th}$ grade whether their math or science teacher treats males and females differently. However, adding this variable as regressor produces an estimate which is not statistically different from zero.

[15] The same specification for male students is not reported since results were not statistically significant.

substitutability between a policy which increases the number of female teachers and another which trains teachers to create a positive learning environment[16].

The above results are robust to the inclusion of whether the student got an A in math or science in 8th grade (Column 5). This reassures us about the potential omission of subject-specific unobservable skills. As expected, if a female student got such a high evaluation in her math or science class in 8th grade, she is less likely to believe that boys are better than girls in math/science.

[16] Also note that, as usual with the FE estimation, the $R^2$ is rather low. However, most of the variability is capture by the student fixed-effects. Indeed, if I run the specification of column 4 as a linear regression with a dummy variable for each female student, the $R^2$ is close to 0.70.

## Table 1: Teacher-student interaction - Girls

| | (1) Basic | (2) Teachers | (3) Parents | (4) Interact | (5) 8 grade |
|---|---|---|---|---|---|
| Boys better than girls in math/science | | | | | |
| Female teacher | -0.060*** | -0.058*** | -0.057*** | -0.127*** | -0.129*** |
| | (0.011) | (0.011) | (0.014) | (0.039) | (0.039) |
| Female teacher*math | | | | -0.007 | -0.011 |
| | | | | (0.028) | (0.028) |
| Boys better in math/science (Teacher) | -0.005 | -0.005 | -0.020 | -0.020 | -0.011 |
| | (0.019) | (0.019) | (0.022) | (0.022) | (0.022) |
| Listens students ideas | -0.041** | -0.038** | -0.049** | -0.099*** | -0.102*** |
| | (0.016) | (0.017) | (0.019) | (0.029) | (0.029) |
| Female teacher*listen | | | | 0.084** | 0.087** |
| | | | | (0.039) | (0.039) |
| More than bachelor | | 0.013 | 0.024* | 0.024* | 0.029** |
| | | (0.011) | (0.013) | (0.013) | (0.014) |
| Bachelor with STEM major | | -0.017 | -0.020 | -0.020 | -0.022 |
| | | (0.012) | (0.015) | (0.014) | (0.015) |
| Experience teaching math/science | | -0.001 | -0.001 | -0.001 | -0.002 |
| | | (0.002) | (0.002) | (0.002) | (0.003) |
| Experience teaching math/science^2 | | 0.000 | 0.000 | 0.000 | 0.000 |
| | | (0.000) | (0.000) | (0.000) | (0.000) |
| Boys better in math/science (Parent) | | | 0.058*** | 0.057*** | 0.054*** |
| | | | (0.019) | (0.019) | (0.019) |
| A in 8th grade math/science | | | | | -0.086*** |
| | | | | | (0.019) |
| Subject math fixed effect | -0.002 | -0.006 | -0.006 | -0.002 | -0.001 |
| | (0.008) | (0.008) | (0.010) | (0.020) | (0.020) |
| Constant | 0.227*** | 0.228*** | 0.226*** | 0.269*** | 0.314*** |
| | (0.016) | (0.022) | (0.026) | (0.033) | (0.036) |
| Observations | 11940 | 11910 | 8380 | 8380 | 8180 |
| Overall R^2 | 0.00499 | 0.00483 | 0.00989 | 0.00889 | 0.00684 |
| Within R^2 | 0.00864 | 0.00920 | 0.01316 | 0.01479 | 0.02379 |

Standard errors in parentheses. SE clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Only female students considered. Source: HSLS09.

## Table 2: Teacher-student interaction - Boys

| | (1) Basic | (2) Teachers | (3) Parents | (4) Interact | (5) 8 grade |
|---|---|---|---|---|---|
| Boys better than girls in math/science | | | | | |
| Female teacher | -0.053*** | -0.054*** | -0.061*** | -0.075*** | -0.080*** |
| | (0.012) | (0.012) | (0.014) | (0.018) | (0.019) |
| Female teacher*math | | | | 0.028 | 0.036 |
| | | | | (0.025) | (0.026) |
| Boys better in math/science (Teacher) | -0.006 | -0.003 | -0.014 | -0.013 | -0.016 |
| | (0.020) | (0.019) | (0.024) | (0.024) | (0.025) |
| Listens students ideas | 0.026 | 0.025 | 0.028 | 0.028 | 0.023 |
| | (0.018) | (0.018) | (0.022) | (0.022) | (0.023) |
| More than bachelor | | 0.015 | 0.013 | 0.012 | 0.010 |
| | | (0.012) | (0.016) | (0.016) | (0.016) |
| Bachelor with STEM major | | 0.011 | 0.018 | 0.019 | 0.019 |
| | | (0.012) | (0.015) | (0.015) | (0.015) |
| Experience teaching math/science | | 0.004** | 0.005* | 0.005* | 0.005* |
| | | (0.002) | (0.003) | (0.003) | (0.003) |
| Experience teaching math/science^2 | | -0.000* | -0.000* | -0.000* | -0.000* |
| | | (0.000) | (0.000) | (0.000) | (0.000) |
| Boys better in math/science (Parent) | | | 0.053*** | 0.053*** | 0.057*** |
| | | | (0.019) | (0.019) | (0.020) |
| A in 8th grade math/science | | | | | 0.036* |
| | | | | | (0.020) |
| Subject math fixed effect | 0.017** | 0.020** | 0.023** | 0.007 | 0.003 |
| | (0.008) | (0.008) | (0.010) | (0.016) | (0.017) |
| Constant | 0.238*** | 0.202*** | 0.187*** | 0.194*** | 0.187*** |
| | (0.018) | (0.023) | (0.028) | (0.029) | (0.031) |
| Observations | 11790 | 11760 | 8040 | 8040 | 7810 |
| Overall R^2 | 0.00192 | 0.00220 | 0.00739 | 0.00790 | 0.01308 |
| Within R^2 | 0.00650 | 0.00849 | 0.01414 | 0.01452 | 0.01663 |

Standard errors in parentheses. SE clustered at school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Only male students considered. Source: HSLS09.

## 6. Empirical Section II: Advanced Math/Science Courses

### 6.1 Econometric Framework

Trying to affect how students compare male and female in math and science is more relevant from a policy perspective if these beliefs actually influence students' decisions during high school and later in life. Therefore, in order to offer an additional motivation for the first empirical section, the aim of this second part of the paper is to investigate whether students' beliefs affect their decision to take advanced math or science classes in high school. Since the scope of the paper is to analyze the lack of women in STEM fields, the estimation focuses on female students' decisions only. Indeed, students in 9th grade are asked if they plan to enroll in an AP/IB calculus or science course.

The identification strategy is similar to the one used in the previous section. The equations of interest are the following:

$$AP_{is} = \beta_1 b_{is} + z'_{is}\beta_2 + x'_{is}\beta_3 + \mu_i + \alpha_s + \varepsilon_{is}\,, \forall s \in \{M, N\}$$

The dependent variable is whether or not student $i$ is planning to enroll in an advance course in subject $s$ while in high school[17]. The key regressor of interest is whether the student $i$ believes that boys are better than girls in subject $s$. However, such enrollment decision depends also on other factors which may be correlated with student beliefs. Therefore, also in this case it is appropriate to add as controls the 9th grade math/science teacher characteristics for student $i$ ($z_{iM}$ and $z_{iS}$), observable and unobservable student fixed-effects ($\mu_i$), and subject fixed-effect ($\alpha_s$).

As in the previous section, we may be worried about omitted variables: the error term could include factors which determine the student's decision to take AP classes in math or science and are correlated with her beliefs. Indeed, according to the (National Science Board, 2014), students in 9th grade are more likely to take courses in science and math if they are from high socioeconomic status categories or if their parents are highly educated. However, the individual fixed-effect allow us to control for such variables. In addition to this, subject-specific family characteristics are taken into account by including parents' beliefs among the regressors ($x_{is}$).

Moreover, the student decision does not only depend on the teacher characteristics, but also on how effective he or she was in engaging the students and offering a high-quality course. In other words, the teacher's characteristics included in the vectors $z_{iM}$ and $z_{iS}$ may not be enough to measure her quality. Thus, the controls include a variable which indicates whether the student enjoyed the course $s$ in 9th grade.

[17] The dependent variable is whether the student was planning to enroll in an advance math or science class during high school when she was interviewed in 9th grade. The follow-up surveys contain the actual number of AP/IB classes taken in high school. The reason why these are not used in the above specification is because of endogeneity issues. Indeed, the actual number of AP/IB courses depends on teachers throughout all years of high school, and they may be correlated with students' beliefs. Since this dataset contains information only about teachers in 9th grade, having dependent variable, students' beliefs and teachers' characteristics all reported in 9th grade solves this issue.

Furthermore, peers can also play an important role in shaping these decisions. Therefore, the above specification includes whether the student wanted to take additional math/science classes because her friends were going to. Another way to control for peer effect is to add the average belief among the $9^{th}$ graders interviewed in the school (excluding the answer for the $i^{th}$ individual). Nevertheless, adding such percentages of interviewed students who believe that boys are better than girls in math or science does not substantially changes the results shown in Table 3. The same is true if such percentages are computed considering only female students.

Finally, individual subject-specific ability may be pivotal. In the next section I have used two strategies to tackle this issue: including among the controls the grades in the most advanced math and science courses in middle school, or a measure of self-efficacy.

### 6.2 Empirical Results

The important result from this section is that, as shown in the first column of Table 3, female students are more than 7 percentage points less likely to enroll in an advance math and science courses in high school if they believe that boys are better than girls in math or science. Therefore, influencing students' beliefs at the beginning of high school can result in female students taking more scientific classes, thus increasing their college readiness.

For later reference, it is important to note that most of the teacher characteristics do not significantly affect students' decisions. In other words, if a teacher creates a positive learning environment by listening to students' ideas does not directly enter into the student decision function. If it influences such choice, it does it indirectly by changing students' beliefs or their satisfaction about the 9$^{th}$ grade courses.

As before, it is possible to check for heterogeneity across subjects in the effect of students' beliefs: contrary to the descriptive statistics highlighted in the previous section, the interaction between beliefs and subject fixed effect is not statistically different from zero (Column 2).

Although I control for student fixed-effects, which includes individual skills used both in math and science courses, as in the previous section one could still be worried that omitted subject-specific skills may drive the decision and be correlated with students' beliefs. One way of dealing with this potential issue is to add whether the student got an A in her 8$^{th}$ grade math or science course (Column 3). The coefficient of student *i*' opinion does not change substantially.

A similar conclusion is reached once a measure of student self-efficacy in math and science is added as a regressor. However, also in this case there are reasons to be worried since the measure is self-reported, thus if a female student believes that girls are better than boys in math, she may also be more confident in her own abilities. In other words, self-efficacy may capture an indirect effect of the regressor of interest. Indeed, in the last column of Table 3 the coefficient of student *i*' belief is lower than in the previous specifications. Despite this, it remains significant and with large magnitude.

## Table 3: Intention to take AP/IB courses in math/science

| | (1) Basic | (2) Interact | (3) 8 grade | (4) Ability |
|---|---|---|---|---|
| Boys better in math/science (Student) | -0.074*** | -0.084*** | -0.087*** | -0.076*** |
| | (0.019) | (0.027) | (0.028) | (0.027) |
| Boys better * Math | | 0.018 | 0.027 | 0.026 |
| | | (0.035) | (0.036) | (0.034) |
| Female teacher | -0.005 | -0.005 | -0.004 | -0.003 |
| | (0.015) | (0.015) | (0.015) | (0.015) |
| Boys better in math/science (Teacher) | -0.019 | -0.019 | -0.023 | -0.021 |
| | (0.022) | (0.022) | (0.022) | (0.022) |
| Listens students ideas | -0.033 | -0.033 | -0.036* | -0.043** |
| | (0.021) | (0.021) | (0.021) | (0.021) |
| More than bachelor | 0.038** | 0.039** | 0.041*** | 0.041*** |
| | (0.015) | (0.015) | (0.015) | (0.015) |
| Bachelor with STEM major | -0.016 | -0.016 | -0.018 | -0.015 |
| | (0.013) | (0.013) | (0.013) | (0.013) |
| Experience teaching math/science | -0.003 | -0.003 | -0.003 | -0.003 |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Experience teaching math/science^2 | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Subject math fixed effect | 0.006 | 0.003 | 0.003 | -0.004 |
| | (0.010) | (0.011) | (0.011) | (0.012) |
| Boys better in math/science (Parent) | -0.038* | -0.038* | -0.036* | -0.030 |
| | (0.021) | (0.021) | (0.021) | (0.021) |
| Friends taking AP | 0.032 | 0.032 | 0.019 | 0.026 |
| | (0.042) | (0.043) | (0.044) | (0.044) |
| Enjoyed 9th grade math/science course | 0.114*** | 0.114*** | 0.115*** | 0.078*** |
| | (0.016) | (0.016) | (0.016) | (0.018) |
| A in 8th grade math/science | | | 0.068*** | |
| | | | (0.021) | |
| Math/science self-efficacy | | | | 0.055*** |
| | | | | (0.009) |
| Constant | 0.395*** | 0.396*** | 0.365*** | 0.426*** |
| | (0.026) | (0.025) | (0.029) | (0.026) |
| Observations | 6130 | 6130 | 6000 | 6050 |
| Overall R^2 | 0.03352 | 0.03388 | 0.08706 | 0.09825 |
| Within R^2 | 0.04523 | 0.04538 | 0.05405 | 0.06614 |

Standard errors in parentheses. SE clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Only female students considered. Source: HSLS09.

## 7. Empirical Section III: STEM Major

### 7.1 OLS

One of the aims of the paper is to reduce the gender gap in STEM field. It has already been shown that high school teachers can change how students compare boys and girls in math and science and that such students' beliefs can affect female students' college readiness. This last section investigates whether the same beliefs can influence female students' major choice.

This is a one-time decision, so it is possible to start with a very simple econometric model

$$STEM_i = \gamma b_i + x'_i\beta + \varepsilon_i$$

The dependent variable is whether student $i$ intends to choose a STEM subject as a major. This question is asked when the student is a freshman in college. The key regressor of interest is whether the student $i$ believes that boys are better than girls in science. Major choice may also depend on other individual observable characteristics ($x_i$) and unobservable ones ($\varepsilon_i$). This model is estimated for female students only since they represent the population of interest.

Most of the students in STEM choose engineering, biology, chemistry or physics, while only a minority opts for mathematics. Therefore, I expect that the main role is played by the student's belief about boys' and girls' abilities in science rather than math.

I may also worry that teachers in the final year of high school may affect both students' belief and major choice. One way to attenuate this issue is to use as regressor student $i$'s belief in $9^{th}$ grade rather than $11^{th}$ grade. This is the result reported in the first column of Table 4.

One reason why this estimation may be biased is because of omitted variables. One possible solution which is usually implemented is to include additional regressors as controls. However, here it is important to carefully choose the appropriate controls. There are some variables which affect the major decision but not students' beliefs, so it is not necessary to include them in a linear model. For instance, labor market conditions such as monetary and non-monetary returns of different major influences students' decision to major or not in a STEM field, but the fact that engineers on average earn more than journalists should not affect girls in $9^{th}$ grade when they compare boys and girls in math and science, so omitting them would not bias our estimator of interest[18]. The appropriate controls are variables which affect both students' beliefs and major choice. Ethnicity, country of origin, geographical regions are examples of such variables. Finally, I am interested in the direct and indirect effects of beliefs on major choice. As a result, I do not want to control for variables which affect major choice but are caused by students' beliefs. Following the previous section, female students who take more AP classes in math and

[18] I may worry that students' opinions are also affected by discrimination against women. In other words, a female student may think that women are discriminated in STEM fields because they are less able than men. This seems implausible (Ceci et al., 2014) and would require different levels of discrimination in STEM and non-STEM fields. Moreover, I have showed in the first empirical section that students' beliefs are not determined by episodes of discrimination.

science while in high school are more likely to choose a STEM major, but they may enroll in more advanced classes because they do not believe that boys are better than girls in math and science, so controlling for it may underestimate the total impact of beliefs on major choice[19].

Given the above discussion, it is appropriate to control for ethnicity, whether the student was born in the US, geographical regions, and whether the school was in an urban area. As usual, peer-effects could also play a role in this context. In order to address this issue, I have controlled for whether the student's best friend in 9$^{th}$ grade was a good student (although the dataset do not include information about the gender of this friend) in the second specification. Family background also play an important role in shaping college decisions and students' opinions. Hence, this specification includes information about household income, parents' education, work experience, beliefs and involvement in various aspect of their daughter's life. This specification includes also some relevant school characteristics such as safety and extra-curricular activities. This is the preferred specification. The coefficient of interest remains negative and significant: if a female student believes that boys are better than girls, she is 5 percentage points less likely to choose her major in a STEM field.

As discussed above, including students' achievements in high school as controls may capture an indirect effect of students' beliefs. Nevertheless, an advantage of this approach is that it allows me to include proxies for abilities as regressors, thus tackling such endogeneity issue. As shown in the last column of Table 4, the high school GPA in STEM classes, the number of STEM credits and the normalized score in the math tests administered during the survey in 9$^{th}$ and 11$^{th}$ grades are highly correlated with the decision of choosing a STEM major. Nevertheless, the coefficient of student's belief remains negative and significant[20].

Another advantage of including these controls is that they reflect students' preferences. In other words, a female student may decide not to pursue a STEM major simply because she does not like math or science. However, these tastes may actually due to her belief: if such a student thinks that boys are better than girls, she may end up actually not enjoying those subject. So preferences would then be an outcome variable and therefore a "bad" control. Furthermore, these latent preferences are actually reflected in the actual choices made by the student during her academic career. Consequently, including the STEM credits and grades reveals and controls for student's preferences.

---

[19] See also (J. D. Angrist & Pischke, 2009) on "Bad Control".

[20] In addition to this, I have also tried to estimate the same model by adding the 9$^{th}$ grade math and science teachers' characteristics as regressors. The estimate of the student's belief does not change substantially, while the impacts of the teachers' gender, belief, class behavior, educational level or college major are not statistically different from zero. Similarly, if I add whether the student believed in 9$^{th}$ grade that boys were better than girls in math, its estimated coefficient is not statistically significant, while the student's belief in science remains negative and significant. On the other hand, if I measure beliefs with a dummy variable equal to one if the student believed in 9$^{th}$ grade that boys are better than girls in science or math, the coefficient is -0.05 and statistically significant. Finally, I have also estimated the same models as in Column 3 Table 4 for male students only. Whether or not a male student believes that boys are better than girls in science does not significantly affect his major choice decision.

As in the previous sections, from an econometric perspective some people may argue that a linear probability model is not the best way to model a limited dependent variable model (major choice) with an endogenous dummy regressor (student belief). However, I follow (Joshua D Angrist, 2001), (J. D. Angrist & Pischke, 2009) and argue that this is appropriate for estimating average causal effects. For the sake of comparison, I have also estimated a Probit model. Results do not change substantially: the marginal effect of the student's belief is almost -0.06 and high significant, thus in line with the OLS estimate.

## Table 4: STEM major intention for female students - OLS

| | (1) Basic | (2) Controls | (3) STEM |
|---|---|---|---|
| Boys better in science (Student) | -0.040*** | -0.050** | -0.063*** |
| | (0.013) | (0.021) | (0.021) |
| Asian | | 0.150*** | 0.081* |
| | | (0.043) | (0.044) |
| Black | | -0.019 | 0.050 |
| | | (0.036) | (0.037) |
| Hispanic | | 0.001 | 0.047* |
| | | (0.026) | (0.027) |
| Other non-white race | | 0.022 | 0.029 |
| | | (0.030) | (0.029) |
| US born | | -0.049 | -0.056 |
| | | (0.039) | (0.036) |
| Best friend has good grades | | -0.043 | -0.062 |
| | | (0.038) | (0.037) |
| HH income 2011 | | -0.001 | -0.004 |
| | | (0.005) | (0.005) |
| HH income 2008 | | 0.003 | 0.003 |
| | | (0.005) | (0.005) |
| Mother highly educated | | 0.006 | -0.019 |
| | | (0.019) | (0.019) |
| Father highly educated | | 0.075*** | 0.041** |
| | | (0.019) | (0.019) |
| Mother in STEM | | 0.045** | 0.046** |
| | | (0.022) | (0.021) |
| Father in STEM | | 0.049** | 0.012 |
| | | (0.025) | (0.024) |
| Parent help homework | | -0.059** | -0.022 |
| | | (0.023) | (0.023) |
| No intellectual activity w/parent | | -0.110** | -0.040 |
| | | (0.049) | (0.057) |
| Boys better in science (Parent) | | -0.004 | -0.001 |
| | | (0.021) | (0.022) |
| Boys better in math (Parent) | | -0.005 | 0.009 |
| | | (0.019) | (0.018) |
| Student feels safe at school | | 0.025 | -0.023 |
| | | (0.035) | (0.033) |
| Algebra 1 remedial course available | | -0.013 | -0.039* |
| | | (0.024) | (0.023) |
| School has math/science fair | | 0.019 | 0.018 |
| | | (0.019) | (0.018) |
| Math/science mentors | | -0.006 | -0.009 |
| | | (0.018) | (0.018) |
| STEM GPA | | | 0.055*** |
| | | | (0.015) |
| STEM credits | | | 0.037*** |
| | | | (0.006) |
| Math test score (9) | | | 0.003** |
| | | | (0.001) |
| Math test score (11) | | | 0.005*** |
| | | | (0.001) |
| Constant | 0.174*** | 0.239*** | -0.561*** |
| | (0.006) | (0.077) | (0.095) |
| Regional dummies | No | Yes | Yes |
| Urban dummies | No | Yes | Yes |
| Observations | 5660 | 2120 | 2020 |
| R^2 | 0.00157 | 0.05507 | 0.15118 |
| AdjR^2 | 0.00139 | 0.04285 | 0.13796 |

Standard errors in parentheses. SE clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Only female students considered. Baseline for ethnicity is white. Source: HSLS09.

### 7.2 IV

As usual with OLS, endogeneity is the main concern. Nevertheless, it should be stress that in this context it is more difficult to explain why the variable of interest, i.e. student's belief, may be endogenous. Indeed, beliefs and major choice may be both affected by demographic characteristics, parents' characteristics and family background, but these factors have been already controlled for. Ability could also be an omitted variable, but the last specification included several proxies for it. Finally, using the lagged value for student's belief should also address the omission of 12$^{th}$ grade teachers' characteristics.

Despite this, I use an IV approach in order to confirm that the OLS results are in line with estimations obtained using more sophisticated econometric techniques. The instruments used in this section are some of the 9$^{th}$ grade teacher's characteristics. The relevance of such instruments has been shown in the first part of the paper: teachers' characteristics are important in affecting students' beliefs. The idea behind such exclusion restriction is inspired by the literature on teacher-added value. Indeed, several researchers have shown that teachers' impact on students' test scores fades out extremely fast in subsequent years ((Jacob, Lefgren, & Sims, 2010), (Rothstein, 2010), (Andrabi, Das, Khwaja, & Zajonc, 2011)). Therefore, conditioning on the exogenous explanatory variables, it is possible to assume that 9$^{th}$ grade teachers do not directly affect students' major choice in college. In other words, college students do not decide whether to choose a major in a STEM field by remembering how their teacher behaved in class when they were 15-years-old. An additional justification for this strategy can be obtained by looking at the results for AP classes in the previous section: teacher characteristics do not directly affect such decision, so it is plausible to assume that a similar patter should occur when students decide their major. In particular, the excluded instruments exploited in this section are whether the 9$^{th}$ grade science teacher listened and valued student's ideas, and whether he or she thought that boys were better than girls in science. One may argue that these variables are correlated with teacher quality, which may in turn affect major choice through its impact on student decisions and performances while in high school. Nevertheless, the specifications below include the high school STEM GPA, the number of STEM credits, the math test scores, and family background information. Therefore, conditioning on these variables, it is possible to claim that these instruments are not correlated with the error term[21].

The model presented in Table 5 is a linear IV model. As shown in the second column, the coefficient of student belief is negative and statistically significant (p-value: 0.051), although larger and less precise than the OLS estimates[22]. Furthermore, having two instruments for one

[21] Furthermore, it is also possible to include indicators for teacher quality as additional controls. Adding whether the 9$^{th}$ grade science teacher has a graduate degree, if the teacher majored in a STEM fields, the number of year (and its squared term) of teaching experience in science does not change the conclusions.

[22] This estimator has been computed using a 2SLS strategy. Therefore, even if the result is robust to heteroscedasticity and clustering at the school level, it is efficient only under homoscedasticity. Using a 2-Step GMM estimation strategy produces estimates that are efficient under arbitrary heteroscedasticity and clustering at the school level. Nevertheless, the results are extremely similar to the 2SLS. In addition to this, it is important to remember that I am using as instruments two dummy

endogenous variable allow to test for the exogeneity of such instruments through a Sargan-Hansen J test. The Hansen p-value is extremely high (0.93), thus it is not possible to reject the null, which increases the confidence in the exogeneity of these instruments[23].

As usual with the IV strategy, one may be concerned about the weakness of the instruments. First, as shown in the first column of Table 5, from the first stage it is clear that whether the teacher listen to student ideas is an important predictor of student beliefs. However, the coefficient of teacher belief is not statistically significant, and the F-test of such excluded instruments is less than 10 (Stock, Wright, & Yogo, 2002). Therefore, in order to dissipate any doubt, I have estimated the same model using a LIML estimation, which is less biased than the 2SLS in case of week instruments (J. D. Angrist & Pischke, 2009). The results shown in the third column are the same of the 2SLS, thus supporting this IV strategy.

To summarize, although the magnitude differs across the above specifications, this section confirms the results obtained with the OLS regression: if a female student believes that boys are better than girls in science, she is less likely to choose a major in a STEM field once in college.

variables, so the much higher IV estimate is mechanically due to the low variability of the instruments. Finally, I have also used a more parsimonious specification than the OLS one. Using the same set of controls do not change the main result, although the estimates are slightly less precise.

[23] Under the assumption that the instruments are valid, it is also possible to test whether the endogenous regressor (student's belief) can be actually treated as exogenous. The p-value is rather low (0.025), thus it is more conservative to actually instrument such regressor.

## Table 5: STEM major intention for female students - IV

| | (1)<br>I Stage | (2)<br>II Stage | (3)<br>LIML |
|---|---|---|---|
| Boys better in science (Student) | | -0.755* | -0.756* |
| | | (0.387) | (0.388) |
| Listens students ideas (Teacher) | -0.075*** | | |
| | (0.029) | | |
| Boys better in science (Teacher) | -0.009 | | |
| | (0.030) | | |
| Asian | -0.058* | 0.095* | 0.095* |
| | (0.031) | (0.053) | (0.053) |
| Black | 0.015 | 0.064 | 0.064 |
| | (0.042) | (0.047) | (0.047) |
| Hispanic | 0.024 | 0.075** | 0.075** |
| | (0.031) | (0.037) | (0.037) |
| Other non-white race | -0.046* | 0.013 | 0.013 |
| | (0.027) | (0.043) | (0.043) |
| US born | -0.067* | -0.070 | -0.070 |
| | (0.040) | (0.055) | (0.055) |
| HH income 2008 | -0.001 | -0.002 | -0.002 |
| | (0.003) | (0.004) | (0.004) |
| Mother highly educated | -0.021 | -0.021 | -0.021 |
| | (0.022) | (0.026) | (0.026) |
| Father highly educated | 0.033* | 0.055* | 0.055* |
| | (0.020) | (0.028) | (0.028) |
| Mother in STEM | 0.006 | 0.035 | 0.035 |
| | (0.022) | (0.029) | (0.029) |
| Father in STEM | 0.001 | 0.024 | 0.024 |
| | (0.024) | (0.031) | (0.031) |
| Parent help homework | -0.007 | -0.019 | -0.019 |
| | (0.022) | (0.029) | (0.029) |
| No intellectual activity w/parent | -0.030 | 0.024 | 0.024 |
| | (0.067) | (0.100) | (0.100) |
| Boys better in science (Parent) | 0.094*** | 0.030 | 0.030 |
| | (0.026) | (0.047) | (0.047) |
| Boys better in math (Parent) | -0.010 | 0.003 | 0.003 |
| | (0.021) | (0.026) | (0.026) |
| Student feels safe at school | -0.017 | -0.082 | -0.082 |
| | (0.047) | (0.055) | (0.055) |
| School has math/science fair | 0.007 | 0.018 | 0.018 |
| | (0.020) | (0.024) | (0.024) |
| Math/science mentors | 0.017 | -0.000 | -0.000 |
| | (0.019) | (0.024) | (0.024) |
| STEM GPA | -0.014 | 0.041** | 0.041** |
| | (0.017) | (0.020) | (0.020) |
| STEM credits | 0.010* | 0.041*** | 0.041*** |
| | (0.006) | (0.009) | (0.009) |
| Math test score (9) | -0.000 | 0.003 | 0.003 |
| | (0.001) | (0.002) | (0.002) |
| Math test score (11) | 0.001 | 0.006*** | 0.006*** |
| | (0.001) | (0.002) | (0.002) |
| Constant | 0.173* | -0.585*** | -0.585*** |
| | (0.096) | (0.124) | (0.124) |
| Regional dummies | Yes | Yes | Yes |
| Urban dummies | Yes | Yes | Yes |
| Observations | 1750 | 1750 | 1750 |
| Hansen J p-value | | 0.92620 | 0.92623 |

Standard errors in parentheses. SE clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Only female students considered. Baseline for ethnicity is white. Source: HSLS09.

## 8. Conclusions

This paper has shown that how female students compare boys and girls in math and science has important implications for their subsequent educational choices. In particular, the empirical analysis has highlighted that female students are more likely to take advance math and science classes in high school and declare a STEM major in college if they believe that boys are not better than girls in math and science. Therefore, policy-maker could increase female workers in STEM fields by affecting these beliefs. In particular, the core of this paper has demonstrated that students' beliefs can be influenced by their 9$^{th}$ grade teachers.

Indeed, female students are less likely to believe that boys are better than girls in math or science when they have a female teacher in those subjects, or when their teacher creates a positive learning environment by listening and promoting students' ideas. Such positive effect is consistent with the role model effect discussed in the previous literature. Furthermore, it is in line with the positive effect of female leaders on girl aspirations and educational attainments (Beaman, Duflo, Pande, & Topalova, 2012).

It is also important to stress that female teachers can also change male students' beliefs and make them less likely to believe that boys are better than girls. This is potentially a groundbreaking result since those boys will continue to interact with women in college, at work, and within their family. As a result, their opinion will help improve the learning and working environment, and it will also have a positive intergenerational effect, thus contributing to break the current vicious circle which pushes women out of STEM fields.

Finally, a note of caution about the magnitude of the aforementioned results: as shown in the descriptive statistics, female students who indeed believe that boys are better than girls are a minority (although a substantial one). In addition to this, the last empirical section demonstrated that it is possible to (considerably) increase women in STEM by changing these beliefs, but it would not be enough to completely fill the gender gap in these fields. In other words, as already discussed, I am not suggesting a panacea, but rather a complementary strategy which could be adopted together with other solutions, such as additional workplace flexibility (Goldin, 2014), in order to speed up the transition to gender equality in science and math.

## Bibliography

Altonji, J. G., Arcidiacono, P., & Maurel, A. (2015). The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. *NBER Working Paper*, *21655*. http://doi.org/10.1017/CBO9781107415324.004

Andrabi, T., Das, J., Khwaja, A. I., & Zajonc, T. (2011). Do Value-Added Estimates Add Value ? Accounting for Learning Dynamics Author. *American Economic Journal: Applied Economics*, *3*(3), 29–54.

Angrist, J. D. (2001). Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors. *Journal of Business & Economic Statistics*, *19*(1), 2–28. http://doi.org/10.1198/07350010152472571

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics : An Empiricist 's Companion*. *Princeton University Press*.

Antecol, H., Ozkan, E., & Serkan, O. (2015). The Effect of Teacher Gender on Student Achievement in Primary School: Evidence from a Randomized Experiment. *Journal of Labor Economics*, *33*(6453), 38. http://doi.org/10.1016/j.econedurev.2013.08.004

Apple. (2015). Inclusion&Diversity. Retrieved from http://www.apple.com/diversity/

Baum, C. F., & Schaffer, M. E. (2012). ivreg2h: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments. Retrieved from http://ideas.repec.org/c/boc/bocode/s457555.html

Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. *Science*, *335*(February), 582–586. Retrieved from www.sciencemag.org/cgi/content/full/335/6068/579/DC1

Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review*, *95*(2), 152–157. http://doi.org/10.1257/000282805774670149

Blow, C. M. (2015, February 2). A Future Segregated by Science? *New York Times*. Retrieved from http://www.nytimes.com/2015/02/02/opinion/charles-blow-a-future-segregated-by-science.html

Bottia, M. C., Stearns, E., Mickelson, R. A., Moller, S., & Valentino, L. (2015). Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM. *Economics of Education Review*, *45*, 14–27. http://doi.org/10.1016/j.econedurev.2015.01.002

Bronson, M. A. (2015). *Degrees Are Forever: Marriage, Educational Investment, and Lifecycle Labor Decisions of Men and Women*.

Cappellari, L., & Jenkins, S. P. (2003). Multivariate probit regression using simulated maximum likelihood. *Stata Journal*, *3*(3), 278–294. http://doi.org/The Stata Journal

Carnevale, A. P., Smith, N., & Melton, M. (2011). *Stem*. Washington, DC. Retrieved from http://cew.georgetown.edu/stem/

Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *The Quarterly Journal of Economics*, *125*(August), 1101–1144. http://doi.org/10.1162/qjec.2010.125.3.1101

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in Academic Science: A Changing Landscape. *Psychological Science in the Public Interes*, *15*(3), 75–141.

http://doi.org/10.1177/1529100614541236

Chen, X. (2013). *STEM Attrition: College Students' Paths Into and Out of STEM Fields*. Washington, DC. Retrieved from http://necs.ed.gov

Coate, S., & Loury, C. G. (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes ? *American Economic Review*, *83*(5), 1220–1240.

Corbett, C., & Hill, C. (2015). *Solving the Equation-The Variables for Women's Success in Engineering and Computing*. Washington, DC: AAUW. Retrieved from http://www.aauw.org/resource/get-the-solving-the-equation-report/

Dee, T. S. (2005). A teacher like me: Does Race, Ethnicity, or Gender Matter? *American Economic Review*, *95*(2), 158–165.

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

Degner, K. M. (2013). Demography as Destiny: The Role of Parental Involvement and Mathematics Course Taking Patterns among 9th Grade Students. *Current Issues in Education*, *16*(3).

Economist, T. (2015). The thinking behind feminist economics. *The Economist*. Retrieved from http://www.economist.com/blogs/economist-explains/2015/10/economist-explains-17

Executive Office of the President. (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics.*

Freedman, D. A., & Sekhon, J. S. (2010). Endogeneity in probit response models. *Political Analysis*, *18*(2), 138–150. http://doi.org/10.1093/pan/mpp037

Friedman-Sokuler, N., & Justman, M. (2016). Gender streaming and prior achievement in high school science and mathematics. *Economics of Education Review*, *53*, 230–253. http://doi.org/http://dx.doi.org/10.1016/j.econedurev.2016.04.004

Fryer Jr, R. G., & Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, *2*(April), 210–240.

Gemici, A., & Wiswall, M. (2014). Evolution of gender differences in post-secondary human capital investments: College majors. *International Economic Review*, *55*(1), 23–56. http://doi.org/10.1111/iere.12040

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teacher Expectations. *Economics of Education Review*, *52*, 209–224. http://doi.org/10.1016/j.econedurev.2016.03.002

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, *104*(4), 1091–1119. http://doi.org/10.1257/aer.104.4.1091

Google. (2015). GoogleDiversity. Retrieved from http://www.google.com/diversity/index.html

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, Gender, and Math. *Science*, *320*(5880), 1164–1165.

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles*, *66*(3-4), 153–166. http://doi.org/10.1007/s11199-011-9996-2

Han, S., & Vytlacil, E. J. (2015). *Identification in a Generalization of Bivariate Probit Models with Endogenous Regressors*.

Hoffmann, F., & Oreopoulos, P. (2009). A Professor Like Me: The Influence of Instructor Gender on College Achievement. *Journal of Human Resources*, *44*(2), 479–494. http://doi.org/10.1353/jhr.2009.0024

Holmlund, H., & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics*, *15*(1), 37–53. http://doi.org/10.1016/j.labeco.2006.12.002

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender Similarities Characterize Math Performance. *Science*, *321*(July), 494–495.

Ingels, S. J., Pratt, D. J., Herget, D., Bryan, M., Fritch, L. B., Ottem, R., … Wilson, D. (2015). *High School Longitudinal Study of 2009 ( HSLS : 09 ) 2013 Update and High School Transcript Data File Documentation* (Vol. NCES). Washington, DC.

Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., … Leinwand, S. (2011). *High School Longitudinal Study of 2009 (HSLS:09). Base-Year Data File Documentation* (Vol. 2011–328). Washington, DC. Retrieved from http://nces.ed.gov/pubsearch

Ingels, S. J., Pratt, D. J., Herget, D. R., Dever, J. A., Fritch, L. B., Ottem, R., … Leinwand, S. (2014). *High School Longitudinal Study of 2009 (HSLS:09) Base Year to First Follow-Up Data File Documentation* (Vol. 2014–361). Washington, DC.

Jackson, C. (2013). *Student Access to Qualified High School Mathematics Teachers: A Multilevel Analysis*. University of Maryland.

Jacob, B. A., Lefgren, L., & Sims, D. P. (2010). The persistence of teacher-induced learning. *Journal of Human Resources*, *45*(4), 915–943. http://doi.org/10.1353/jhr.2010.0029

Kramer, N. C., Karacora, B., Lucas, G., Dehghani, M., Ruther, G., & Gratch, J. (2016). Closing the gender gap in STEM with friendly male instructors? on the effects of rapport behavior and gender of a virtual agent in an instructional interaction. *Computers and Education*, *99*, 1–13. http://doi.org/10.1016/j.compedu.2016.04.002

Leef, G. (2014). True Or False: America Desperately Needs More STEM Workers. *Forbes*, *June 6*. Retrieved from http://www.forbes.com/sites/georgeleef/2014/06/06/true-or-false-america-desperately-needs-more-stem-workers/#728ef1a57225

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 23–34. http://doi.org/10.1081/E-EWS

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67–80. http://doi.org/10.1080/07350015.2012.643126

Mechtenberg, L. (2009). Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages. *The Review of Economic Studies*, *76*(4), 1431–1459.

Mourifie, I., & Meango, R. (2014). A note on the identification in two equations probit model with dummy endogenous regressor. *Economics Letters*, *125*(3), 360–363. http://doi.org/10.1016/j.econlet.2014.10.006

Muralidharan, K., & Sheth, K. (2016). Bridging Education Gender Gaps in Developing Countries: The Role of Female Teachers. *Journal of Human Resources*, *51*(2), 269–297. Retrieved from http://www.nber.org/papers/w19341

National Center for Education. (2013). The Nation's Report Card: Trends in Academic Progress 2012.

*National Center for Education Statistics*, *NCES 2013-*. http://doi.org/10.1002/yd.20075

National Science Board. (2014). *Science & Engineering Indicators 2014*. *NSB 14-01*. Arlington, VA. Retrieved from http://www.nsf.gov/statistics/seind14/index.cfm/etc/pdf.htm

National Science Foundation. (2015). *Women, Minorities, and Persons with Disabilities in Science and Engineering*. *Special Report NSF 15-311*. Arlington, VA. Retrieved from http://www.nsf.gov/statistics/wmpd/

NCES. (2015). *Digest of Education Statistics, 2013*. Washington, D.C.

NCES. (2016). Distance Learning Dataset Training. Retrieved August 31, 2016, from http://nces.ed.gov/training/datauser/

OECD. (2014). PISA 2012 Results in Focus. *OECD Publishing*, *PISA*, 1–44. Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

OECD. (2015). The ABC of Gender Equality in Education. *OECD Publishing*, *PISA*. http://doi.org/10.1787/9789264229945-en

Page, S. E. (2008). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press. Retrieved from http://press.princeton.edu/titles/8757.html

Paredes, V. (2014). A teacher like me or a student like me? Role model versus teacher bias effect. *Economics of Education Review*, *39*, 38–49. http://doi.org/10.1016/j.econedurev.2013.12.001

Patterson, B., & Kobrin, J. (2012). *The SAT and SAT Subject Tests: Discrepant Scores and Incremental Validity*. *College Board*. Retrieved from http://research.collegeboard.org/publications/sat-and-sat-subject-tests-discrepant-scores-and-incremental-validity

Peracchi, F., & Rossetti, C. (2012). Heterogeneity in health responses and anchoring vignettes. *Empirical Economics*, *42*(2), 513–538. http://doi.org/10.1007/s00181-011-0530-8

Pollack, E. (2013, October 3). Why Are There Still So Few Women in Science? *New York Times*. Retrieved from http://www.nytimes.com/2013/10/06/magazine/why-are-there-still-so-few-women-in-science.html?_r=0

Pollack, E. (2015, October 10). College Major Choice and the Gender Gap. *New York Times*. Retrieved from http://www.nytimes.com/2015/10/11/opinion/sunday/what-really-keeps-women-out-of-tech.html

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*(6), 901–910. http://doi.org/10.1016/j.econedurev.2010.07.009

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(12), 4403–8. http://doi.org/10.1073/pnas.1314788111

Rosen, L. (2013). The Truth Hurts: The STEM Crisis Is Not a Myth. *Huffington Post*, *Nov 11*. Retrieved from http://www.huffingtonpost.com/linda-rosen/the-truth-hurts-the-stem-_b_3900575.html

Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *The Quarterly Journal of Economics*, *125*(1), 175–214. http://doi.org/10.1162/qjec.2010.125.1.175

Sabia, J. J. (2007). The Effect of Body Weight on Adolescent Academic Performance. *Southern Economic Journal*, *73*(4), 871–900.

Sabot, R., & Wakeman-Linn, J. (1991). Grade Inflation and Course Choice. *Journal of Economic Perspectives*, *5*(1), 159–170.

Schneeweis, N., & Zweimüller, M. (2012). Girls, girls, girls: Gender composition and female school choice. *Economics of Education Review*, *31*(4), 482–500. http://doi.org/10.1016/j.econedurev.2011.11.002

Sikdar, S. (2015). On efforts in teams with stereotypes. *Economics Letters*, *137*, 203–207. http://doi.org/10.1016/j.econlet.2015.10.032

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What Are We Weighting For ? *Journal of Human Resources*, *50*(2), 301–316.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *The American Psychologist*, *60*(9), 950–958. http://doi.org/10.1037/0003-066X.60.9.950

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, *20*(4), 518–529. http://doi.org/10.1198/073500102288618658

Turner, S. E., & Bowen, W. G. (1999). Choice of Major: The Changing (Unchanging) Gender Gap. *Industrial and Labor Relations Review*, *52*(2), 289–313. http://doi.org/10.1017/CBO9781107415324.004

Varma, R. (2010). Why So Few Women Enroll in Computing? Gender and Ethnic differences in students' perception. *Computer Science Education*, *20*(4), 301–316.

Vella, F. (1994). Gender Roles and Human Capital Investment: The Relationship between Traditional Attitudes and Female Labour Market Performance. *Economica*, *61*(242).

Wagstaff, I. R. (2014). *Predicting 9th Graders' Science Self-efficacy and STEM Career Intent: A Multilevel Approach. North Carolina State University*. North Carolina State University.

Weinberger, C. J., & Leggon, C. B. (2004). Just Ask! Why Surveyed Women Did Not Pursue IT Courses or Careers. *Ieee Technology And Society Magazine*, *Spring*, 28–35.

Wheeling, K. (2015, November). The brains of men and women aren't really that different, study finds. *Science*. http://doi.org/10.1126/science.aad7499

Wilde, J. (2000). Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters*, *69*(3), 309–312. http://doi.org/10.1111/j.1467-9574.2004.00122.x

Williams, M. (2014). Building a More Diverse Facebook. Retrieved from http://newsroom.fb.com/news/2014/06/building-a-more-diverse-facebook/

Winters, M. A., Haight, R. C., Swaim, T. T., & Pickering, K. A. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review*, *34*, 69–75. http://doi.org/10.1016/j.econedurev.2013.01.007

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (II Edition). MIT Press.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(October), 686–688. http://doi.org/10.1017/CBO9781107415324.004

Zafar, B. (2013). College Major Choice and the Gender Gap. *Journal of Human Resources*, *48*(3), 545–595. http://doi.org/10.1353/jhr.2013.0022

# Appendix

## A1. Summary Statistics

**Table A1: Summary statistics**

| Variable | Obs | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| STEM major | 11,580 | 0.256 | 0.436 | 0 | 1 |
| Boys better in science (9th grader) | 20,660 | 0.202 | 0.402 | 0 | 1 |
| Boys better in math (9th grader) | 20,720 | 0.197 | 0.398 | 0 | 1 |
| Boys better in science (11th grader) | 19,980 | 0.240 | 0.427 | 0 | 1 |
| Boys better in math (11th grader) | 20,010 | 0.274 | 0.446 | 0 | 1 |
| Plan to take AP/IB math | 15,120 | 0.392 | 0.488 | 0 | 1 |
| Plan to take AP/IB science | 15,370 | 0.388 | 0.487 | 0 | 1 |
| Female student | 25,150 | 0.489 | 0.500 | 0 | 1 |
| Asian | 25,150 | 0.083 | 0.276 | 0 | 1 |
| Black | 25,150 | 0.105 | 0.307 | 0 | 1 |
| Hispanic | 25,150 | 0.159 | 0.366 | 0 | 1 |
| White | 25,150 | 0.487 | 0.500 | 0 | 1 |
| Other race | 25,150 | 0.089 | 0.284 | 0 | 1 |
| Born in the US | 16,180 | 0.921 | 0.269 | 0 | 1 |
| 9th grader's best friend has good grades | 20,950 | 0.873 | 0.333 | 0 | 1 |
| Northeast | 25,150 | 0.158 | 0.365 | 0 | 1 |
| Midwest | 25,150 | 0.265 | 0.441 | 0 | 1 |
| South | 25,150 | 0.405 | 0.491 | 0 | 1 |
| West | 25,150 | 0.172 | 0.377 | 0 | 1 |
| City | 25,150 | 0.287 | 0.452 | 0 | 1 |
| Suburb | 25,150 | 0.365 | 0.481 | 0 | 1 |
| Town | 25,150 | 0.116 | 0.320 | 0 | 1 |
| Rural | 25,150 | 0.233 | 0.423 | 0 | 1 |
| 2011 Household Income | 21,160 | 4.622 | 3.049 | 1 | 13 |
| 2008 Household Income | 16,950 | 4.630 | 3.048 | 1 | 13 |
| Mother highly educated | 15,750 | 0.510 | 0.500 | 0 | 1 |
| Father highly educated | 13,200 | 0.492 | 0.500 | 0 | 1 |
| Mother in STEM | 15,200 | 0.179 | 0.383 | 0 | 1 |
| Father in STEM | 13,200 | 0.144 | 0.351 | 0 | 1 |
| Parent helps with homework | 15,890 | 0.777 | 0.416 | 0 | 1 |
| No intellectual activity w/parent | 15,620 | 0.021 | 0.142 | 0 | 1 |
| Boys better in science (Parent) | 14,900 | 0.215 | 0.411 | 0 | 1 |
| Boys better in math (Parent) | 14,930 | 0.300 | 0.458 | 0 | 1 |
| STEM GPA | 21,820 | 2.421 | 0.929 | 0 | 4 |

| | | | | | |
|---|---|---|---|---|---|
| **STEM credits** | 21,920 | 7.468 | 2.689 | 0 | 20 |
| **Math test score (9th grader)** | 21,440 | 51.110 | 10.078 | 24.018 | 82.188 |
| **Math test score (11th grader)** | 20,590 | 51.504 | 10.154 | 22.238 | 84.905 |
| **A in 8th grade science** | 20,540 | 0.412 | 0.492 | 0 | 1 |
| **A in 8th grade math** | 20,820 | 0.371 | 0.483 | 0 | 1 |
| **Friends taking more science courses** | 20,060 | 0.056 | 0.231 | 0 | 1 |
| **Friends taking more math courses** | 20,420 | 0.054 | 0.226 | 0 | 1 |
| **Student feels safe at school** | 21,130 | 0.913 | 0.282 | 0 | 1 |
| **Algebra 1 remedial course available** | 16,310 | 0.832 | 0.374 | 0 | 1 |
| **School has math/science fair** | 21,610 | 0.394 | 0.489 | 0 | 1 |
| **Math/science mentors** | 21,610 | 0.360 | 0.480 | 0 | 1 |
| **Female teacher (Math)** | 17,880 | 0.608 | 0.488 | 0 | 1 |
| **More than bachelor (Math teacher)** | 17,870 | 0.507 | 0.500 | 0 | 1 |
| **Bachelor with STEM major (Math teacher)** | 17,840 | 0.402 | 0.490 | 0 | 1 |
| **Experience teaching math (Math teacher)** | 17,820 | 10.328 | 9.009 | 1 | 50 |
| **Boys better in math (Math teacher)** | 16,000 | 0.106 | 0.307 | 0 | 1 |
| **Listens students ideas (Math teacher)** | 18,970 | 0.856 | 0.351 | 0 | 1 |
| **Female teacher (Science)** | 16,260 | 0.563 | 0.496 | 0 | 1 |
| **More than bachelor (Science teacher)** | 16,260 | 0.566 | 0.496 | 0 | 1 |
| **Bachelor with STEM major (Science teacher)** | 16,250 | 0.578 | 0.494 | 0 | 1 |
| **Experience teaching science (Science teacher)** | 16,210 | 10.868 | 9.151 | 1 | 48 |
| **Boys better in science (Science teacher)** | 14,510 | 0.090 | 0.286 | 0 | 1 |
| **Listens students ideas (Science teacher)** | 17,500 | 0.860 | 0.347 | 0 | 1 |
| **School offers AP courses on-site** | 23,630 | 0.897799 | 0.30292 | 0 | 1 |

Note: these statistics are reported for the whole sample without weights and may be different for the actual sample used in some of the econometric specifications. Also note that all sample size numbers are rounded to the nearest 10 for security reasons.

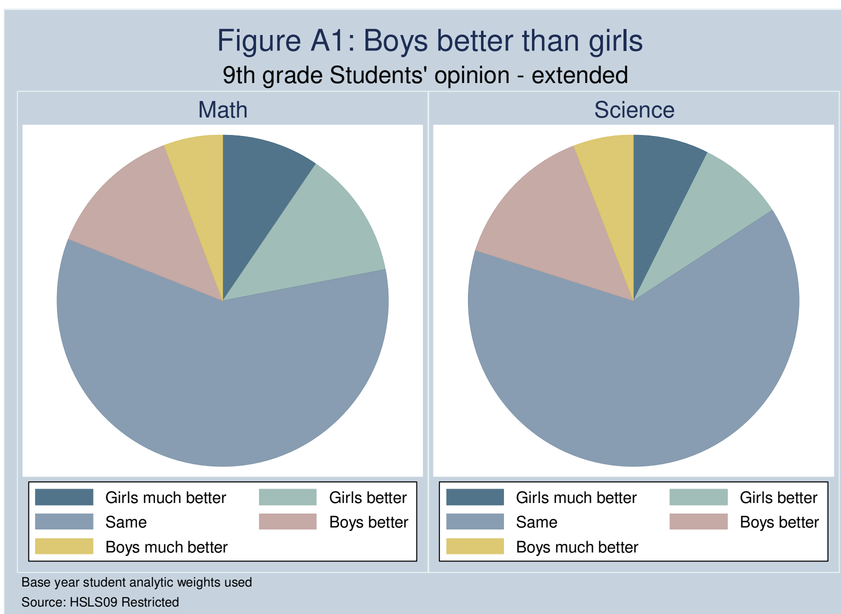
## A2. Variable Description

This appendix contains the detailed description of all the variables used in the empirical models which are not self-explanatory.

*Respondents' beliefs*. Students in 9th and 11th grades, as well as parents, 9th grade math and science teachers are asked to compare boys and girls in math and science. The original question is the following:

> How would you compare males and females in each of the following subjects? (math or science)
> 1. Females are much better
> 2. Females are somewhat better
> 3. Females and males are the same
> 4. Males are somewhat better
> 5. Males are much better

As shown in Figure A1, most of the students believed that males and females are the same in math and science when they are in 9th grade. The same is true in 11th grade, even if the proportion of students who believed that girls are better decreased both in math and science. On the other hand, the percentage of those who believed that boys and girls are the same was constant over time in science, while it decreased in math. The first three options and the last two have been aggregated in the empirical analysis based on the assumption that the key treatment is whether or not the individual believes that boys are better than girls. This has been done not only to simplify the analysis, but also to compare more easily answers across individuals. Indeed, if one imagines that each individual had a continuous latent variable which measure how much boys are better than girls, it is likely that each respondent will have a different threshold which allows him or her to say that "Males are much better" instead of "Males are somewhat better" (Peracchi & Rossetti, 2012). Instead, it is easier to justify a common threshold for the treatment dummy. For instance, one may imagine that all individuals compare math abilities of the men and women they know,



and declare that males are better if more men than women are above average.

*AP/IB math/science course*. Students in 9th grade were asked if they were planning to enroll in an Advanced Placement (AP) calculus course or an International Baccalaureate (IB) calculus course. This dummy variable takes value one if the student answered “Yes” to one of the two questions, i.e. whether he/she planned to enroll in an AP or IB calculus course, while it takes value zero it the student answered “No” or “Don’t know what AP/IB calculus is”. In other words, it is assumed that if student will not take an advance class if she does not even know of their existence. The cases in which the student reported “Haven’t decided yet”, as well as the observations for which the question was not asked/answered, were considered as missing. The aim of the paper is to analyze how to increase the proportion of female student who intend to enroll in an advance math class, so the appropriate comparison group are those students who do not plan to attend such classes, while it is not clear whether those undecided will eventually select such courses. The same method applies to the questions regarding AP or IB science courses.

*STEM major*. In the Summer-Fall of 2013 it was asked to all respondents who were attending postsecondary classes which field of study or program they were considering. Note that most of the respondents were freshman in college at that time. This dummy variable takes value one if the student was considering a STEM field as a major, zero if he/she was considering a different major, missing if he/she was undecided or if the question was not asked or answered. The sample of female students has around 11,360 observations. Among these, 4,980 students were choosing a non-STEM major, 1000 students intended to declare a STEM major, 550 were undecided, while to 2,610 individuals the question was not asked (mainly because they were not attending college), 2,220 missing (sample numbers rounded to the nearest ten for security reason).

*Classroom environment*. Students were asked how much they agree with the following statement (Fall 2009): “Your math teacher values and listens to students’ ideas”. The indicator variable takes value one if the student strongly agreed or agreed, while it takes value zero if the student disagreed or strongly disagreed, and it is missing if the question was not asked or answered. Students were also reminded that none of their teachers and principal were allowed to see their answers. The same question was asked for the science teacher.

*Teacher’s highest education*. An indicator variable has been constructed which takes value zero if the 9th grade teacher held an Associate’s or Bachelor’s degree, while it takes value one if the highest degree earned by the teacher was a Master’s degree, an Educational Specialist diploma, a Ph.D., M.D., law degree or similar.

*Teacher’s major*. This indicator variable takes value one if the teacher’s major for his or her BA/BS was in a STEM field. By far, the most common STEM majors were Mathematics and

Statistics among math teachers, Biology and Physics among science teachers.

*Teacher experience*. The math teacher reported how many years, including the current school year, he or she had taught high school (grade 9-12) math. A specular question was asked to the science teacher. Following the approach stressed in the educational literature, nonlinearities are taken into consideration by including experience as a polynomial of grade two in the empirical models.

*8^th^ grade*. The students were asked to report their final grades in the most advanced 8th grade math and science courses. The derived dummy variables are equal to one if the student obtained an A in those courses, while it is equal to zero if the student got a B, C, D or lower. If the courses were not graded or the question was not asked/answered, the variables are reported as missing.

*Friends taking AP*. This variable indicates whether the 9th grade student planned to take more math or science courses during high school because his or her friends were going to do the same.

*Enjoy 9th grade course*. Students were asked how much they agree with the following statement about their 9th grade math or science course (Fall 2009): "You are enjoying this class very much". The indicator variable takes value one if the student strongly agreed or agreed, while it takes value zero if the student disagreed or strongly disagreed, and it is missing if the question was not asked or answered.

*Self-efficacy*. This variable is a scale of the respondent's math self-efficacy. Higher values represent higher math self-efficacy. This variable had been created by NCES through principal components factor analysis (weighted by base year student analytical weights) and standardized to a mean of 0 and standard deviation of 1. The inputs to this scale were:

- Student reported being confident that she could do an excellent job on tests in the Fall 2009 math course,
- Student reported being certain that she could understand the most difficult material presented in the textbook used in the Fall 2009 math course,
- Student reported being certain that she could master the skills being taught in the Fall 2009 math course,
- Student reported being confident that she could do an excellent job on assignments in the Fall 2009 math course.

Only respondents who provided a full set of responses were assigned a scale value. If the student indicated that he or she was not taking a math class in the Fall of 2009, this variable has been set to missing. A similar procedure had been applied to derive the measure of self-efficacy in science.

*Ethnicity*. The indicator variable *Asian* is equal to one if the respondent's race is Asian non-Hispanic, *Black* refers to Black/African-American non-Hispanic, *Hispanic* to Hispanic individuals with or without their race specified, *White* to White non-Hispanic, while *Other* includes American Indians, Alaska Natives, Native Hawaiian, Pacific Islanders, and individuals with more than one race. White is used as comparison group.

*US Born*. This indicator variable is equal to one if the parent interviewed in 9th grade reported that his/her son or daughter was born in the United States, Puerto Rico or another U.S. territory. If the student was born in another country, the variable takes value zero.

*Best friend has good grades*. This indicator variable is equal to one if the 9th grade student reported that his/her best friend had good grades.

*Household Income in 2008 and 2011*. These are two categorical variables which indicates the student's family income from all sources in 2008 and 2011, as reported by the parent questionnaire respondent. If missing from the parent questionnaire, these variables were statistically imputed. The income categories are the following:

1. Family income less than or equal to $15,000
2. Family income > $15,000 and <= $35,000
3. Family income > $35,000 and <= $55,000
4. Family income > $55,000 and <= $75,000
5. Family income > $75,000 and <= $95,000
6. Family income > $95,000 and <= $115,000
7. Family income > $115,000 and <= $135,000
8. Family income > $135,000 and <= $155,000
9. Family income > $155,000 and <=$175,000
10. Family income > $175,000 and <= $195,000
11. Family income > $195,000 and <= $215,000
12. Family income > $215,000 and <= $235,000
13. Family income > $235,000

*Mother/Father highly educated*. This indicator variable is equal to one if the student's biological, adoptive or step mother had an Associate's degree, Bachelor's degree or higher, while it is zero in case of high-school diploma, GED or less than high school. The same categorization applies for the father.

*Mother/Father in STEM*. This indicator variable is equal to one if the student's biological, adoptive or step mother's most recent occupation was in a STEM field. The following 2-digit ONET codes has been considered as STEM: Computer Science and Mathematics; Architecture and Engineering; Life, Physical, and Social Science; Healthcare Practitioners and Technical Occupations; Healthcare Support Occupations. The same categorization applies for the father.

*Parent help homework*. This indicator variable is equal to zero if the student's parent had never helped his/her son or daughter with homework in 9$^{th}$ grade, one if he/she had helped the 9$^{th}$ grader less than once a week or more often.

*No intellectual activity w/parent*. This indicator variable is equal to one if the student's parent or another family member had never done any of the following activities in the previous 12 months:

- Visiting a zoo, planetarium, natural history museum, transportation museum, or a similar museum.
- Working or playing on a computer together.
- Building or fixing something such as a vehicle or appliance.
- Attending a school science fair.
- Helping 9$^{th}$ grader with a school science fair project.
- Discussing a program or article about math, science, or technology.
- Visiting a library.
- Going to a play, concert, or other live show.

*Student feels safe at school*. This indicator variable is equal to one if the 9$^{th}$ grader strongly agreed or agreed with the sentence "You feel safe at this school", while it is equal to zero if he/she disagreed or strongly disagreed.

*Algebra 1 remedial course available*. This indicator variable is equal to one if the 9$^{th}$ grader's math teacher judged the availability of tutoring or other remedial assistance for students who were struggling in Algebra 1 as good or excellent, while it is equal to zero in case of a poor/fair evaluation.

*School has math/science fair*. This indicator variable is equal to one if the 9$^{th}$ grader's principal reported that his/her school held school-wide math or science fairs, workshops, or competitions; zero if not.

*Math/science mentors*. This indicator variable is equal to one if the 9$^{th}$ grader's principal reported that his/her school paired students with mentors in math or science; zero if not.

*STEM GPA*. This variable is extracted from the student's high school transcripts and it contains the student's GPA for STEM courses.

*STEM credits*. This variable is extracted from the student's high school transcripts and it contains the student's total Carnegie credits in STEM courses.

*Math test scores*. These mathematical standardized scores provide a norm-referenced measurement of achievement, that is, an estimate of achievement relative to the population (Fall 2009 9th graders) as a whole. It provides information on status compared to peers. This feature is the main difference from the IRT-estimated percent-correct scores, which represent status with respect to achievement on a particular criterion set of test items. Such mathematical assessment focused on algebra skills, reasoning, and problem solving. It was developed specifically for the HSLS:09 and was administered to students in $9^{th}$ grade and $11^{th}$ grade. See Chapter 2 in (Ingels et al., 2011) and (Ingels et al., 2014) for more information about the test and the derivation of the normalized scores.

*Regional indicators*. Four indicator variables were created to identify the geographical region of the $9^{th}$ grade student's school. The macro-regions considered here are Northeast (comparison group), Midwest, South, and West.

*Urban indicators*. Four indicator variables were created to identify the locale (urbanicity) of the $9^{th}$ grade student's school. The school neighborhood could be described as City (comparison group), Suburb, Town, or Rural.

## A3. Sensitivity Analysis for Empirical Section I: Student-Teacher Interaction

### A3.1 OLS

In order to analyze the student-teacher interaction, it may be useful to look at the OLS estimates as well. In particular, the equations of interest are the following (for boys and girls separately):

$$b_{Mi} = z'_{Mi}\beta_{1M} + x'_{Mi}\beta_{2M} + \varepsilon_{Mi}$$

$$b_{Ni} = z'_{Ni}\beta_{1N} + x'_{Ni}\beta_{2N} + \varepsilon_{Nit}$$

As in the main equations, the dependent variable is whether or not student $i$ thinks that boys are better than girls in math ($b_{Mi}$) or science ($b_{Ni}$). The relevant variables are the math/science teacher characteristics for student $i$ ($z_{Mi}$ and $z_{Ni}$). The same controls of Tables 1 and 2 have been used in this analysis.

The single-equation estimates are shown in Table A2 for boys (columns 1 and 2) and girls (columns 3 and 4) for math and science respectively. The teacher's gender is significantly correlated with students' beliefs both in math and science and for boys and girls. It is also worth mentioning that there is some heterogeneity in the magnitude across subjects: the coefficient of female teacher is larger in science than in math. Similarly, whether the teacher listens and values students' ideas matters for girls, especially in math, but not for boys. Consistently with the FE analysis, teachers' beliefs do not seem to be relevant in this context. Overall, except for the between-subject differences, the OLS results are qualitatively similar to the main ones presented in Tables 1 and 2. For the sake of completeness, I have also estimated a Probit model using the same regressors. The results from marginal effects are very similar to the OLS estimates.

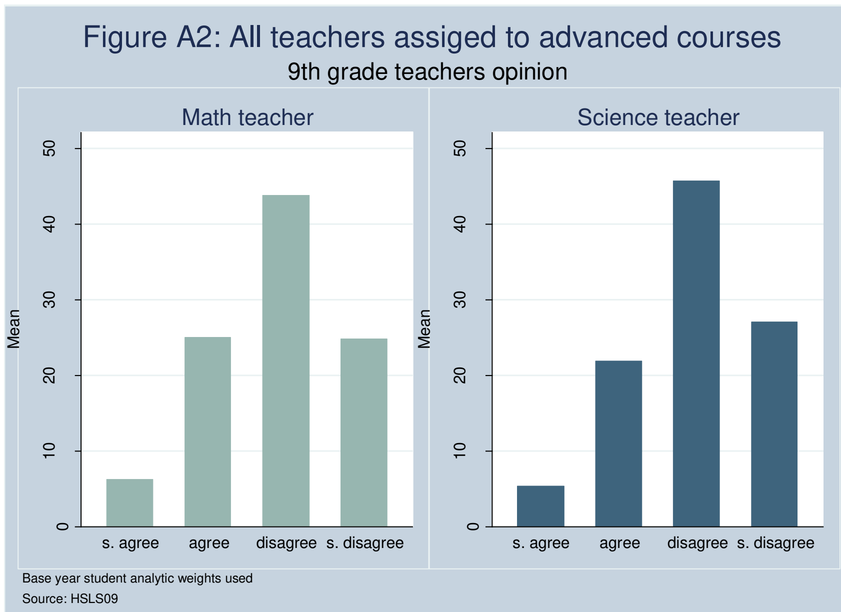
## Table A2: Teacher-student interaction - OLS

| | (1) Boys Math | (2) Boys Science | (3) Girls Math | (4) Girls Science |
|---|---|---|---|---|
| Female teacher (Math Teacher) | -0.018 | | -0.028** | |
| | (0.014) | | (0.012) | |
| Female teacher (Science Teacher) | | -0.066*** | | -0.040*** |
| | | (0.014) | | (0.012) |
| Boys better in math (Math Teacher) | 0.049** | | -0.018 | |
| | (0.024) | | (0.018) | |
| Boys better in science (Science Teacher) | | 0.039 | | -0.019 |
| | | (0.024) | | (0.019) |
| Listens students ideas (Math Teacher) | 0.007 | | -0.075*** | |
| | (0.019) | | (0.019) | |
| Listens students ideas (Science Teacher) | | -0.007 | | -0.042** |
| | | (0.021) | | (0.018) |
| More than bachelor (Math Teacher) | -0.020 | | 0.006 | |
| | (0.014) | | (0.011) | |
| More than bachelor (Science Teacher) | | 0.000 | | 0.040*** |
| | | (0.015) | | (0.012) |
| Bachelor with STEM major (Math Teacher) | 0.030* | | 0.018 | |
| | (0.015) | | (0.012) | |
| Bachelor with STEM major (Science Teacher) | | -0.000 | | -0.010 |
| | | (0.014) | | (0.012) |
| Experience teaching math (Math Teacher) | 0.004 | | 0.000 | |
| | (0.003) | | (0.002) | |
| Experience teaching math^2 (Math Teacher) | -0.000 | | -0.000 | |
| | (0.000) | | (0.000) | |
| Experience teaching science (Science Teacher) | | 0.003 | | -0.004* |
| | | (0.003) | | (0.002) |
| Experience teaching science^2 (Science Teacher) | | -0.000 | | 0.000* |
| | | (0.000) | | (0.000) |
| Boys better in math (Parent) | 0.084*** | | 0.068*** | |
| | (0.014) | | (0.013) | |
| Boys better in science (Parent) | | 0.058*** | | 0.056*** |
| | | (0.017) | | (0.016) |
| Constant | 0.187*** | 0.244*** | 0.205*** | 0.209*** |
| | (0.024) | (0.027) | (0.024) | (0.025) |
| Observations | 4230 | 3810 | 4430 | 3950 |
| R^2 | 0.01416 | 0.01076 | 0.01444 | 0.01135 |
| AdjR^2 | 0.01230 | 0.00868 | 0.01266 | 0.00935 |

Standard errors in parentheses. SE clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: HSLS09

### A3.2 Sensitivity Analysis

One concern expressed in the literature is about the sorting of students into classroom. For instance, low ability female students may be systematically assigned to female teachers. Therefore, teachers' characteristics would no longer be exogenous. This would not be an issue if such sorting were based on student observable characteristics which are controlled for. Similarly, if the sorting mechanism were the same for math and science teacher, it would be taken into account by the fixed-effects. But if sorting were based on student unobservables and different across subjects, it would raise an endogeneity problem. One way to dissipate this concern is to verify how teachers are actually assigned between classes. In the HSLS:09, 9th grade math and science teachers are asked to what extend they agree or disagree with the statement "All or most [math/science] teachers are assigned at least one section of advanced courses". As shown in Figure A2, the answers are similar between the two groups. The same conclusion can be reached by comparing how much math and science teachers agree with the following statement: "Advanced courses are assigned to teachers with the strongest [math/science] background". Almost 60% of the math teachers and 64% of the science teachers agree or strongly agree with it. Therefore, this evidence suggest that the sorting mechanism is in most cases the same across subject, thus it has been already taken into account in the previous section by including the student fixed-effects.




In a specular way, it is likely that students are sorted similarly in math and science classes: if for instance a student or a parent has a preference for female teachers, this would be true in both math and science classes, so it would be included in the student fixed-effect. In addition to this, information about the placement policy in the schools is provided in the HSLS:09 by the school counselors. Specifically, they are asked about the importance of student/parent choice for 9th grade science/math class. Their answers are similar for the two subjects[24], thus supporting our

[24] According to the counselors' answers, students/parent have a small or no role in the 9th grade placement in 25% of the schools for math and 21% for science. This question was asked only in case all 9th graders were not placed in the same math/science class.

conclusion. In addition to this, the main specification already controls for parental beliefs, so it already takes into account different sorting due to parents' decisions.

As an additional robustness check, I have also estimated the same equations as in the third columns of Tables 1 and 2 by restricting the sample to schools in which all 9$^{th}$ grade students are assigned to the same math and science courses, so the selection process is limited by construction. Although the sample size is substantially reduced, the coefficient of teacher gender remains statistically significant both for boys and girls.

Another concern may be reverse causality: teachers' beliefs may depend on students' beliefs. However, under the assumption that teachers form their beliefs over time and that such beliefs are a function of all previous students, it is possible to claim that after a certain number of years those beliefs should be stable and the effect of an additional student on the teacher's beliefs should be marginal. Therefore, I estimated the same specifications as the ones in the main section (Tables 1 and 2) for boys and girls separately only for teachers who have at least 3 years of experience teaching high school math/science. The results are similar to the previous estimates: teachers' gender affects male and female students' beliefs, female students are also influenced by whether the teacher listens and values students' ideas.

There may also be heterogeneous effects: teachers may have a different impact on top students and on low-achievers. The HSLS:09 includes a math test scores in 9$^{th}$ grade. Therefore, I estimated the same specifications as the ones in the main section (Tables 2 and 3) for boys and girls separately only for students whose standardized theta score was above the overall median[25]. The results for the high-achievers are qualitatively similar to those of the whole sample.

A potential issue is generated by the survey collection activities: students were not interviewed all at the same time. Therefore, even if all students were affected by the new teachers, some of them could have been more influenced since they were interviewed at the end of the semester, thus they had spent more time with their teachers. Nevertheless, excluding students interviewed between September and October 2009 does not substantially change the results.

This time variation in the interview process could potentially be exploited as a robustness check. Indeed, students interviewed later has been exposed to the "treatment", i.e. to a different teacher, for a longer period. Therefore, we could use students interviewed at the beginning of the semester as a control group. In a regression model for students' beliefs, this is equivalent to add as control an indicator variable equal to one if the student was interviewed earlier in the Fall of 2009, and to interact such variable with the teacher's key characteristics (gender, belief, behavior). However, such interactions are not statistically significant both for the math and science regression. This is consistent with the idea that the teacher may have an immediate

[25] Note that this approach is not ideal since a student's performance in the math test may depend on his or her beliefs and characteristics, as well as his or her teachers' characteristics. However, this is the best proxy available for genetic math ability.

effect, or a much stronger effects in the first few weeks, thus the two groups would not differ substantially.

Another issue is about spillover effects: math teachers may influence students' beliefs in science, and science teachers may influence students' beliefs in math. In order to explain this case, let's focus for simplicity only on the basic specification. There would be two equations:

$$b_{Mi} = z'_{Mi}\beta_M + z'_{Ni}\gamma_N + \mu_i + \varepsilon_{Mi}$$

$$b_{Ni} = z'_{Ni}\beta_N + z'_{Mi}\gamma_M + \mu_i + \varepsilon_{Nit}$$

So by taking the difference across subjects would result in

$$(b_{Mi} - b_{Ni}) = z'_{Mi}(\beta_M - \gamma_M) + z'_{Ni}(\gamma_N - \beta_N) + (\varepsilon_{Mi} - \varepsilon_{Ni})$$

It seems plausible to assume that the coefficients $\beta_M$ and $\gamma_M$ have the same sign. In other words, if a female math teacher leads fewer students to believe that boys are better than girls in math, it should also lead fewer students to believe that boys are better than girls in science. However, the magnitude should be larger in math, i.e. $\beta_M > \gamma_M$. Therefore, in this case the estimates would capture the difference between the two channels, thus underestimating the direct impact of a female math teacher on student beliefs in math. A similar reasoning applies to the other teacher characteristics, as well as to the science teacher. Indeed, this underestimation may be the reason behind the fact that the coefficients of teachers' beliefs are not statistically different from zero in Tables 1 and 2. If I explicitly estimate the above FD equation, I obtain results qualitatively similar to the previous section: female teacher has a significant and substantial impact on students' beliefs both for boys and girls[26].

Another variation that I could have exploited is the change in student's belief between 11$^{th}$ and 9$^{th}$ grade. However, only information about the 9$^{th}$ grade math and science teachers is available, very little information is reported about the 11$^{th}$ grade teachers. In particular, I do not observe if the 11$^{th}$ grade math and science teachers are female, how they compare boys and girls in math and science, and if they listen and value students' ideas. Therefore, regressing changes in math individual beliefs on 9$^{th}$ grade teacher characteristics would just estimate whether the impact of such teacher on the student's belief in 9$^{th}$ grade is different from the impact on the same belief in 11$^{th}$ grade. In other words, even if I run such regressions and I found that whether the teacher listen to students has a positive and statistically significant coefficient for female students (which is the actual estimation result), this is just a confirmation of the previous results. Indeed, a positive learning environment makes girls less likely to think that boys are better than girls in math, and the contemporaneous effect is larger than the effect on beliefs 2 years later. On the

[26] The estimated equations include also parents' beliefs as additional controls. The main difference compared with the estimates in Tables 1 and 2 is that the impact of teacher behavior is not statistically significant in all specification. Note also that the main econometric difference between this FD equation and the main FE model is that in Tables 1 and 2 only female teacher was interacted with the subject fixed effect, while here I introduce more heterogeneity since I look at both math and science teacher, so it is as if I interacted all regressors for the subject fixed-effect.

other hand, the coefficient of female math teacher is not statistically significant, which may be due to the fact that the gender of the 9th grade math teacher has an impact on beliefs in both years of similar magnitude.

In addition to this, I have also estimated the same model adding students' beliefs in 9th grade as a regressor. In this way, I can test whether 9th grade teachers have a direct effect on the change in students' beliefs during high school in addition to the simultaneous effect on beliefs in 9th grade. The results indicate that female science teachers not only lead fewer female students to believe that boys are not better than girls in science, but they also make them more likely to change beliefs over time. In other words, a female student with a female science teacher is more likely to increase her confidence in women's abilities in science between 9th and 11th grade. This holds also when I regress the change over time in the student's science beliefs on both the math and science teachers' characteristics. On the other hand, all the other key indicators, including the gender of the math teacher, are not statistically significant. Note that these results are consistent with the selected exclusion restrictions in Paragraph 7.2.

There are two other omitted variables which may endanger our estimation strategy. One is previous teachers. For instance, math and science teachers in 8th grade may affect students' beliefs. Nevertheless, if a previous teacher was the same in the math and science courses, this would be included in the student fixed-effect. Moreover, it is difficult to see how the 8th grade teachers may be correlated with the regressors of interest, i.e. 9th grade teachers' gender, beliefs and opinions. Put differently, although previous teachers may influence students' opinions and assignment of students to 9th grade courses, they should not create an omitted variable bias since they do not usually affect the assignments of teachers in 9th grade across courses.

The other omitted variable is peer effect: if the best student in the class is a girl, or if the student's best friend is a girl with good grades in math and science, this would probably influence the student's comparison of boys and girls in math and science. This dataset does not contain this kind of information. Therefore, this is a clear limitation of this study. However, if the best friend is not a classmate, this should not be correlated with 9th grade teacher characteristics. Furthermore, if the best friend does not change across subject (and he/she has similar grades in math and science), this is included in the student fixed-effect. The same is true for the top student if she is a girl in both the math and science classes. In addition to this, even if the top student is a girl only in one class, this may be due to the teacher. Therefore, controlling for such peer effect may not be desirable since it would underestimate the teacher's impact by excluding such an indirect effect.

## A4. Sensitivity Analysis for Empirical Section II: Advanced Math/Science Courses

So far it has been assumed that the key treatment effect for a female student is to switch from thinking that boys are better than girls in math or science to thinking that boys are not better (or vice versa). In other words, female students in the treatment group think that boys are better than girls in math or science, while those in the control group think either that boys and girls are the same, or that girls are better. Therefore, one way to check for heterogeneity in this model is to split the control group. This is equivalent to adding as additional regressor in the above model another indicator variable equal to one if the $9^{th}$ grader believed that girls are better than boys in math or science, zero otherwise. By doing so, I obtain that if a female student thinks that boys are better, she is almost 6 percentage points less likely to enroll in an advance math or science class while in high school (compare with -0.074 in the first column of Table 4). On the other hand, the coefficient of the new regressor is positive (0.047) with a p-value of 0.06. In addition to this, I have also tried to use a trichotomous variable: zero if the student believes that girls are better, one if she believes that boys and girls are the same, two if she believes that boys are better. The estimated coefficient is -0.05 and it is highly significant.

It is also important to remember that there are still schools in which AP courses are not offered. In fact, around 10% of the schools in the sample do not offer any AP course on-site. In particular, 23% do not offer Calculus AB, while 34% do not offer Biology even if they are both popular AP courses in the US. Nevertheless, results similar to the main section are obtained by estimating the same equations as the ones in Table 3 but only for students in schools which offer AP classes on-site. The same can be said if the same specifications of Table 3 are augmented by including whether the school offer AP/IB courses in Calculus or Biology as regressors.

## A5. Sensitivity Analysis for Empirical Section III: STEM Major

In order to take into account the endogeneity of students' beliefs in their college major choice, a different approach than the traditional IV follows (Lewbel, 2012) and uses heteroscedasticity to estimate a triangular model. In other words, without exclusion restrictions, the relevant system is the following:

$$b_i = x'_i\beta_1 + \nu_i$$

$$STEM_i = \gamma b_i + x'_i\beta_2 + \varepsilon_i$$

Where the notation is the same of the OLS section (Paragraph 7.1). Identification is obtained by assuming that

$$Cov(x_i, \varepsilon_i^2) \neq 0$$

$$Cov(x_i, \nu_i^2) \neq 0$$

$$Cov(z_i, \varepsilon_i \nu_i) = 0$$

Where $z_i$ is observed and it can be a subset of $x_i$. The first two assumptions are common and state that there is heteroscedasticity in the error term. On the other hand, the last one is the key assumption and needs to be carefully considered. This is a heteroscedasticity covariate restriction and, as discussed in the original paper, it is satisfied immediately if the two error terms are conditionally independent. Moreover, even if the error terms are correlated, the assumptions can hold in the classical measurement error framework or when a common unobserved factor is included in both equations. The latter is the most relevant in this case since such unobserved factor could be individual ability. Lastly, since here both $STEM_i$ and $b_i$ are binary indicators, it should be pointed out that the last assumption would not hold if $z_i$ coincided with $x_i$. Therefore, only a proper subset of the exogenous regressors have been included in zi. The estimate coefficient of student's belief is consistent with the linear IV estimate showed in Table 5, although the magnitude is closer to the OLS estimate (-0.27 and p-value of 0.66)[27].

Another approach imposes normality on the distribution of the error terms. It is then possible to estimate the following bivariate probit model.

$$b_i^* = x'_{1i}\beta_1 + \nu_i$$

$$STEM_i^* = \gamma b_i + x'_{2i}\beta_2 + \varepsilon_i$$

[27] This is consistent with (Sabia, 2007). This estimate has been obtained by including in $z_i$ all the demographic and geographical indicators, and by adding whether the science teacher listens to students' ideas and his/her belief to the list of excluded instruments. These estimates have been obtained using the Stata command *ivreg2h* (Baum & Schaffer, 2012). Similar estimates has been obtained with only one exclusion restriction (teacher listens to students) and the same $z_i$. In addition to this, similar estimates have also been obtained without any exclusion restriction and by including in $z_i$ all the exogenous regressors except the school characteristics.

$$b_i = 1 \; if b_i^* > 0, zero \; otherwise$$

$$STEM_i = 1 \; if STEM_i^* > 0, zero \; otherwise$$

$$\begin{pmatrix} \nu_i \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Where the star superscript indicates the latent variable. Note that $x_{1i}$ and $x_{2i}$ can include different regressors. Indeed, (Mourifie & Meango, 2014) showed that a bivariate Probit regression with a dummy endogenous regressor, without heteroscedasticity and without exclusion restrictions is in general only partially identified, thus contradicting (Wilde, 2000). Furthermore, (Freedman & Sekhon, 2010) argued against using a two-step correction in these kind of models. However, they showed that under certain conditions the model is identified. Similarly, (Han & Vytlacil, 2015) showed that an exclusion restriction is sufficient (but not necessary) for identifications when the two equations have common exogenous covariates. It is then possible to use the same exclusion restrictions of the linear IV (Table 5), i.e. science teacher's behavior and belief. Using the resulting estimates, the marginal effect of student's beliefs on the (marginal) probability of choosing a STEM major is -0.39 with the 95% confidence interval [-0.61; -0.16].

A natural extension consists in estimating a trivariate probit model (Cappellari & Jenkins, 2003) where one equation refers to the student's beliefs in science, one to the student's beliefs in math, and one to the STEM major choice. The latter, similarly to the bivariate probit model just discussed, includes both beliefs as regressors. As in the other models, the exclusion restriction for the first equation is satisfied by using $9^{th}$ grade science teacher's belief and behavior. In addition to this, the second equation (math belief) has $9^{th}$ grade math teacher's belief and behavior as additional regressors. The estimated coefficient can then be used to estimate the marginal effect of student's science beliefs on the (marginal) probability of choosing a STEM major. Such effect is negative, and then magnitude is more similar to the OLS estimate than the bivariate probit one. A similar result is obtained if both teachers' behaviors and beliefs are included in the two equations as exclusion restrictions, or if only teacher' behavior is used as exclusion restriction.

In addition to this, I have also tried alternative instruments for female students' belief in the linear IV model. In particular, I have attempted to use the average beliefs among the interviewed students in the same school. Indeed, $9^{th}$ graders are likely to be affected by their peers. As expected, the first stage shows a strong correlation between student's and peers' beliefs. Moreover, there is no clear direct connection (after having controlled for the observable determinants) between major choice in college and what the fellow students in high school believed in $9^{th}$ grade. Therefore, peers' average belief appears to be a valid instrument. Unfortunately, the estimates in the second stage are extremely imprecise, so it is not possible in this case to claim that the (LATE) coefficient of student's belief is statistically different from zero.

A final worry is related, again, to a potential omitted variable bias. Indeed, the HSLS:09 does not include any information on students' beliefs before 9th grade. However, this should not be an issue in the second part of the analysis since the student's beliefs in 8th grade affect the decision to take advance classes in math and science, as well as the college major choice, only through her beliefs later in high school. In other words, they do not have a direct effect on these dependent variables. As far as the first section is concern, beliefs in 8th grade do have a direct effect on beliefs in 9th grade and, since they may differ between math and science, they may not be included in the between-subject student fixed effect. Nevertheless, the parents' beliefs and the student's performances in 8th grade should offer an approximation for such beliefs. Despite these controls, I cannot exclude that if the students had different beliefs in math and science in 8th grade, this may have affected her selection of math and science teachers in 9th grade. In other words, a limitation of this study is that it is not possible to verify whether the sorting mechanism is influenced by the students' beliefs in middle school.

**Latest CeRP Working Papers**

| | | |
|---|---|---|
| N° 165/16 | Dario Sansone | Teacher Characteristics, Student Beliefs and the Gender Gap in STEM Fields |
| N° 164/16 | Noemi Oggero<br>Mariacristina Rossi | Does financial literacy of parents matter for the educational outcome of children? |
| N° 163/16 | Sara Burrone<br>Elsa Fornero<br>Mariacristina Rossi | Retirement Decisions, Eligibility and Financial Literacy |
| N° 162/16 | Elsa Fornero<br>Mariacristina Rossi<br>Dario Sansone | “Four Bright Coins Shining at Me”. Financial Education in Childhood, Financial Confidence in Adulthood |
| N° 161/16 | Riccardo Calcagno<br>Flavia Coda Moscarola<br>Elsa Fornero | Too busy to stay at work. How willing are Italian workers “to pay” to anticipate their retirement? |
| N° 160/16 | Elisa Luciano<br>Antonella Tolomeo | Information effects in longevity-linked vs purely financial portfolios |
| N° 159/16 | Margherita Borella<br>Michele Belloni | Self-Employment in Italy: the Role of Social Security Wealth |
| N° 158/16 | Claudio Morana | Macroeconomic and Financial Effects of Oil Price Shocks: Evidence for the Euro Area |
| N° 157/16 | Riccardo Calcagno<br>Maela Giofré<br>Maria Cesira Urzì-Brancati | To trust is good, but to control is better: how do investors discipline financial advisors’ activity |
| N° 156/16 | Elisa Luciano<br>Mariacristina Rossi<br>Dario Sansone | Financial Inclusion and Life Insurance Demand; Evidence from Italian households |
| N° 155/16 | Claudio Morana | The US$/€ exchange rate: Structural modeling and forecasting during the recent financial crises |
| N° 154/15 | Vincenzo Andrietti | Auto-enrollment, Matching, and Participation in 401(k) Plans |
| N° 153/15 | Donatella Baiardi<br>Claudio Morana | Financial deepening and income distribution inequality in the euro area |
| N° 152/15 | Ewa Gałecka-Burdziak<br>Marek Góra | The impact of easy and early access to old-age benefits on exits from the labour market: a macro-micro analysis |
| N° 151/15 | Margherita Borella<br>Flavia Coda Moscarola | The 2011 Pension Reform in Italy and its Effects on Current and Future Retirees |
| N° 150/15 | Anna Lo Prete | Labour market institutions and household consumption insurance within OECD countries |
| N° 149/15 | Flavia Coda Moscarola<br>Ugo Colombino<br>Francesco Figari<br>Marilena Locatelli | Shifting Taxes from Labour to Property. A Simulation under Labour Market Equilibrium |
| N° 148/15 | Flavia Coda Moscarola<br>Elsa Fornero<br>Steinar Strøm | Absenteeism, Pension Reforms and Grandmothers |

**The full series is available at:** http://www.cerp.carloalberto.org/category/publications/working-papers/