Do professors really perpetuate the gender gap in science? Evidence from a natural experiment in a French higher education institution

Thomas Breda, Son Thierry Ly

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2012 – 13

Do professors really perpetuate the gender gap in science? Evidence from a natural experiment in a French higher education institution

Thomas Breda
Son Thierry Ly

JEL Codes: I23, J16

Keywords: Discrimination ; Gender Stereotypes ; Natural Experiment ; Sex and Science

PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ECOLE NORMALE SUPÉRIEURE – INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE

Do professors really perpetuate the gender gap in science? Evidence from a natural experiment in a French higher education institution[1]

Thomas Breda[2] and Son Thierry Ly[3]

Abstract

Stereotypes, role models played by teachers and social norms are known to push girls to choose humanities rather than science. Do professors directly contribute to this strong selection by discriminating more against girls in more scientific subjects? Using the entrance exam of a French higher education institution (the *Ecole Normale Supérieure*) as a natural experiment, we show the opposite: discrimination goes in favor of females in more male-connoted subjects (e.g. math, philosophy) and in favor of males in more female-connoted subjects (e.g. literature, biology), inducing a rebalancing of sex ratios between science and humanities majors. We identify discrimination by systematic differences in students' scores between oral tests (non-blind toward gender) and anonymous written tests (blind toward gender). By making comparisons of these oral/written scores differences between different subjects for a given student, we are able to control both for a student's ability in each subject and for her overall ability at oral exams. The mechanisms likely to drive this positive discrimination toward the minority gender are also discussed.

JEL codes: I23, J16

Keywords: discrimination, gender stereotypes, natural experiment, sex and science

[1] We would like to thank Philippe Askenazy, Mathilde Gaini, Julien Grenet, Eric Maurin, Thomas Piketty and Helge Thorsen for their helpful comments on this manuscript and Ecole Normale Supérieure for allowing us access to their entrance exam records.

[2] Centre for Economic Performance, London School of Economics.

[3] Ecole normale supérieure, Paris School of Economics.

1. Introduction

Why are there so few girls in science? Although gender differences have disappeared or evolved in favour of girls in many educational outcomes such as college enrolment, male and female students are still strongly segregated across majors. Females compose only 25% of the science, technology, engineering, and math workforce (National Science Foundation 2006) whereas for instance less than one third of new college students in humanities at Yale in 2006 were males[4]. Understanding the origin of these discrepancies is important from an economic perspective: gender differences in entry into science careers accounts for a significant part of the gender pay differential among college graduates (Brown & Cororan 1997; Weinberger 1999) and may also reduce aggregate productivity (Weinberger 1998).

The reasons for the underrepresentation of women in science have been debated by several academic papers, government reports as well as pro-women lobbies. Some important contributions have been made in the literature. We first know that gender differences in math and science test scores are very small and have lowered during the past twenty years. Weinberger (2001) has shown that these small gender differences in abilities do not explain the gender gap in science careers: conditional on proxies for ability, women are still between 50% and 70% less likely than men to complete a degree in science, technology, engineering, or math (Weinberger 2001). Many studies have also established that professors may serve as role models in higher education and that professors' gender strongly affect female college students' attainment and their likelihood to major in science (Canes & Rosen 1995; Rothstein 1999; Gardecki & Neumark 1998; Bettinger & Long 2005; Hoffman & Oreopoulos 2009; Carrell, Page & West 2010). Finally, the gender differences in preferences documented by the experimental literature, such as the gender differences in risk aversion, taste for competition or altruism, have also been put forward as candidate explanations for the gender gap in science majors' enrolment.

By looking at the determinants of students' educational and career choice, the literature on gender gaps across college majors has mostly focused on the supply side. But the equilibrium share of female finally observed in science (and in each other major) is at the intersection between the supply and the demand for female in the field. However, only little is known on the exact role played by the demand side in shaping the observed gender gap in science

[4] See http://www.yaledailynews.com/news/2006/apr/27/gender-gap-in-majors-persists/ or http://oir.yale.edu/yale-factsheet

careers. Do science professors want girls in their course, and more broadly, in their field? If not, women may rationally shy away from science if they know that they are likely to be discriminated in science careers. Women may also be implicitly or explicitly (discrimination) driven away from science majors by professors.

In this paper, we study the direct contribution of the demand side to gender segregation across majors in higher education. Using a unique dataset on the entrance exam of a French top higher education institution, the *Ecole Normale Supérieure* (ENS), for years 2004 to 2009, we show that girls are favored by the exam's jury in the more masculine areas (such as math) whereas boys are favored by the exam's jury in the more feminine areas (such as biology or literature). Our results thus indicate that the demand for students in each major is biased in favor of the minority gender, which induces a reduction of gender segregation across majors. Although the magnitude of the bias is large in many subjects, its direction may be opposite for the same candidate with regard to the subject. As a consequence, the overall effect lowers only slightly the huge gender gaps by major already induced by the relative supply of males and females candidates in each major of the ENS entrance exam.

Our paper relates to the literature on gender stereotypes. It has been known for long that gender stereotypes affect teachers' perceptions (Dusek & Joseph 1983; Madon et al. 1998), which in turn affect the way they evaluate their pupils (Bernard 1979), and the way children perceive their own ability (Tiedemann 2000). A typical gender stereotype is that boys excel in math and science and girls excel in other subjects (Deaux & LaFrance 1998). On the basis of such a stereotype, girls may be encouraged to pursue traditional female studies instead of math or science. Such a behavior has been documented by indirect evidence based on subjective questionnaires answered by parents of first grade students (Carr, Jessup & Fuller 1999) or by PhD holding students (Rowsey 1997), as well as by psychological tests of perception (Glick, Wilk & Perreault 1995). From these studies and the literature on gender stereotypes, we might certainly have expected a stronger discrimination against females in more male-connoted subject areas. By showing the opposite, our results indicate that professors' evaluations are not directly driven by simplistic stereotypes such as "girls are not good in science". This discrepancy between what we find and what was suggested by earlier work may be explained by differences in methodology. Previous literature linking gender stereotypes to gender segregation across subjects was largely based on declared perceptions. However, our paper is (to our knowledge) the first to use a natural experiment to explicitly study the effect of gender discrimination on gender segregation. Interestingly, a similar discrepancy can be found in a slightly different context: using a natural experiment similar to ours, Lavy (2008) shows that high school teachers in Israel systematically discriminate in favor of girls. His results also go against the general view according to which gender stereotypes should harm girls at school. In this paper, we push Lavy's work one step further by explicitly identifying stereotypes (e.g. "girls are not good in science", "girls are good in humanities") and by showing that discrimination actually goes against these precise stereotypes.

To identify the existence of gender discrimination, we use a natural experiment based on the design of the ENS entrance exam. The ENS candidates have to take both a blind written test (their gender is not known by the professor who grades the test) and a non-blind oral test in several subjects. As long as the scores obtained by the candidates at written and oral tests are comparable, i.e., as long as they measure the same skills and cognitive abilities, the blind written score can be used as the counterfactual measure to the non-blind oral score, which may be affected ("treated") by discrimination. This identification framework is similar to that used by Lavy (2008) in a context that relates closely to ours, by Goldin and Rouse (2000) and by Blank (1991). We use the difference between the males' and females' gaps between the blind and the non-blind test scores as a measure of a potential gender bias in a given subject. We control for various observable characteristics of the ENS candidates, for reversion to the mean and for the initial ability of the candidates. In particular, we show that the distributions of the scores of male and female candidates are remarkably similar at written tests in all subjects, whereas sizeable differences appear at oral tests.

In the context of this study, our proposed identification strategy has two potential shortcomings. The first one is that female handwriting may differ from male handwriting. Females can thus be detected at written tests. As a consequence, written tests may not be completely blind. We argue that this problem is not likely to be important. Discussions with several professors or teachers that have already graded written tests suggest that a candidate's gender is not so easy to detect with certainty at written tests. Grading a supposedly female-handwritten test is also very different from facing the physical presence of a female or male candidate at an oral exam. More importantly, the fact that written tests are not perfectly blind with respect to gender can only lead us to underestimate gender discrimination: in the extreme case where gender is perfectly detectable at written tests and affect the jury similarly in both written and oral tests, we should not find any difference between males' and females' gaps between the oral and written tests.

The second shortcoming is that written and oral tests do not measure exactly the same skills. Characteristics such as oral expression, appearance, self-confidence or shyness are likely to affect the candidates' scores at oral tests a lot more than their scores at written tests. If there are systematic differences between males and females regarding these characteristics, the double differences estimates cannot be interpreted as discrimination. However, the nature of this study allows us to overcome this problem to a large extent. Indeed, the design of the natural experiment we use gives the opportunity to implement two distinct strategies, both relying on "triple-differences".

First, we compare males and females not only between oral and written tests, but also between majors. The dataset includes results of students coming from 5 different majors (or tracks) with very different share of female candidates: Math-Physics, Physics-Chemistry, Biology-Geology, Social sciences and Humanities. For example, females have represented 11.6% of the 250 students admitted from the Math-Physics track and 58.5% of the 453 students admitted from the Humanities track over the period 2004-2009. An important feature of this study is to document a premium for female candidates between the oral and written tests that varies positively with the "degree of masculinity" of the candidates' major. In the same way, we show that if the admission had been decided only on the basis of the written tests (i.e. ignoring the oral tests), there would have been 4.1 percentage points less girls in the Math-Physics track and 2.6 percentage points more girls in the Humanities track. For these results to be explained by gender differences in oral tests skills, one has to argue that girls' oral abilities are higher than boys' in the Math-Physics track whereas they are lower in the Humanities track. These kinds of hypothesis that may bias our results are already quite complex and probably less likely to be true. As illustrated by this example, our goal is not to identify the absolute value of the discrimination that affects the candidates' scores at oral tests. It is rather to study how such discrimination varies depending on the candidates' major. Adding this third dimension – the degree of feminization of a given major –improves our identification because it allows us to get rid of all the difference in abilities that are uniform across majors.

However, there is already a lot of gender segregation across majors among the ENS candidates. This is likely to reflect a strong selection process of the candidates. We may not expect that the roughly 10% of females in the Math-Physics track are identical to the roughly

60% of females in the Humanities track in terms of their oral abilities. The few women that have decided to major in Math-Physics despite strong social norms against such a choice may have particular preferences and unobserved characteristics (to the econometrician). They may be for example especially self-confident, which in turn may affect their performance at oral tests. Comparing the oral premium for girls in the different majors does not tackle this issue. We thus also carry out within-candidate between-subjects comparisons. A nice feature of our data is indeed that each candidate in a given track is observed both at written and oral tests in several subjects (5 in average). It is thus possible to investigate how the oral/written test gap evolves across subjects for a same candidate. By differentiating the oral/written test gap between subjects, we control for the general ability of the candidates at oral tests. This individual fixed effects strategy strengthens our previous results: the oral premium for a given girl in a given track is higher on average in the most masculine subjects (computer sciences, mathematics, physics) as compared to the most feminine ones (foreign languages, literature, biology), and the opposite is true for boys.

Finally, we discuss the different mechanisms that are likely to drive our results. First, the ENS and its jury members may implement a conscious affirmative action towards the minority gender in each major. The fact that we find strong differences across subjects within a given track suggests that this is not a good explanation. We also test for the presence of a discontinuity in the distribution of average scores at the admission threshold using the technique developed by McCrary (2007). Indeed, if our results reflect a kind of quota policy implemented (non-officially) by the ENS, we may expect that some females are pushed up *ex post* above the admission threshold, inducing a detectable discontinuity in the average scores' distribution. We do not find evidence of such a discontinuity.

However, three other mechanisms may explain our results. The first is that the professors that grade the candidates have an unconscious taste for the gender in minority and practice a pure preference-based discrimination: math professors are just happier when they have the unusual occasion to interview a female candidate whereas the same is true for literature professors with respect to male candidates. The oral premium towards the minority gender would in that case reflect a kind of taste for diversity.

In contrast, the second plausible mechanism is directly linked to students' abilities. Paradoxically, professors may rationally favor girls in science even if they have negative stereotypes against their abilities. This may happen if they value not only the actual performance but also other qualities such as self-investment and perseverance: if girls are in

average less able than boys in science, they probably had to exhibit more efforts and motivation to reach the same performance level. As the ability to provide lots of efforts and self-investment may be important for their long-term achievements (ENS students are meant to pursue research careers), professors may want to value these attributes.
The last plausible mechanism is information-based and is concerned with statistical discrimination. If the professors grading the oral tests do not perfectly observe the candidates' abilities and believe that the few girls' candidates in the Math-Physics track are positively selected, they may use such a prior when grading the oral tests and consequently favor these girls. Having a positive *a priori* towards the gender in minority might be rational for the jury: when a female (resp. male) has chosen a major that is socially strongly connoted for males (resp. females), she/he is probably highly motivated and feels good enough to manage in an environment that may be adverse (see Fryer, 2007, for a formal presentation of this "belief flipping" mechanism). After having presented our main results, we discuss how these three mechanisms are likely to drive them.

The remainder of this paper is organized as follows. Section 2 describes our data. Section 3 presents our empirical specifications. Our main results come in section 4. Section 5 discusses three potential underlying mechanisms that are likely to drive our results.

2. Data

As our empirical specifications strongly depend on the design of our natural experiment, we start by describing our data and the functioning of the ENS entrance exams.

Ecole Normale Supérieure of Paris entrance exams

The French higher education system is said to be particularly selective: after high school, the best students can enter into a very difficult 2 years preparatory school that prepares them for the entrance exams of selective universities called *Grandes Ecoles*. About 10% of high school graduates choose this way and are selected into a specific major: the main historic ones are Mathematics-Physics, Physics-Chemistry, Biology-Geology, Humanities, Social Sciences. The major in which a student is involved in the preparatory school determines the set of *Grandes Ecoles* in which she may candidate, as well as the set of subjects on which she will be tested. These *Grandes Ecoles* are divided into 4 groups: 215 *Ecoles d'Ingénieur* for scientific and technical studies (the most famous is called *Ecole Polytechnique*), a few hundred *Ecoles de Commerce* for management and business studies, a few hundred Schools for studies in biology, agronomy or veterinary, and three *Ecole Normale Supérieure (ENS).* The number of available places in each *Grande Ecole* is predefined and limited, implying that the *Grandes Ecoles* entrance exams are in fact contests.

The three ENS are aimed to prepare students for high-level teaching and academic careers positions (about 80% of their students eventually do a PhD). The ENS of Paris on which this study focuses is the most prestigious of them and the yearly entrance exams are designed to select the best performing students through a set of very demanding tests. The ENS are also the only *Grandes Ecoles* to be generalist: they accept students from the five historical preparatory schools' majors. As a consequence, the entrance exams for the ENS of Paris are divided into 5 groups that we call "tracks". Candidates from a given major in preparatory schools apply in the track that corresponds to this major and they compete only with other students from the same major. They are tested in a set of subjects that is specific to their track (see Appendix tables A1 and A2). However, a nice feature of the ENS entrance exams is that many subjects are common across tracks, although the tests' precise content remains track-specific[5]. Importantly, both the difficulty of the tests and the jury of the ENS entrance exams remain track and subject specific. This means for example the math test in the Math-Physics track is more difficult and graded by a different jury than the math test in the Social-Sciences track.

The overall structure of the exam is the same in all tracks. Students take a first "eligibility" step of written tests (about 3500 candidates from all majors every year) and all candidates from a given major are then ranked according to a weighted average of all written test scores. The best-ranked students are declared eligible for the second step (the threshold is major-specific with a total of about 500 eligible students). This second "admission" step consists in oral[6] tests on the same subjects[7]. Finally, eligible candidates of each major are ranked according to a weighted average of all written and oral test scores and the best ones are admitted in the ENS. The admission threshold is again major-specific and defined by law (see Table 1 for the yearly average number of eligible and admitted candidates from each major).

[5] As will appear later on, our empirical strategy relies extensively on the fact that the ENS accepts students from different majors and that some subjects are common across majors.
[6] Eligible candidates at scientific tracks also have to take written tests at the admission step.
[7] Teachers never know the grades obtained by the student at the written tests.

We only focus on the roughly 500 students that are eligible for the oral exams each year. We have data for years 2004 to 2009, giving us the universe of the 3068 eligible candidates that took both the written and oral steps in one of the five main tracks of the ENS entrance exam (table 1). 36% of these eligible candidates were finally accepted in the ENS[8]. 40% of both the eligible and finally admitted candidates are girls. However, the proportion of female candidates varies dramatically across majors (see table 1). For example, girls only account for 9% of the candidates in the Math-Physics track whereas they account for 64% of the candidates in Humanities. Interestingly, the proportion of girls among admitted candidates is higher than their proportion among eligible candidates only in the most scientific tracks. Our data also include some individual characteristics for candidates of years 2006-2009 only. We know their social background, the preparatory school they come from, if they got their *Baccalaureat* (the national exam at the end of high-school) with honors and if they were a repeater in their preparatory school[9]. There are some significant gender differences concerning these variables: females are more likely to have obtained their *Baccalaureat* with high honors in most tracks and they are more likely to come from a high social background in the Humanities track (see Appendix table A3). To control for the potential biases that these discrepancies could induce, we include these variables in some of our empirical specifications.

In each track, eligible candidates take a given set of written and oral exams in various subjects (see table 2). Unfortunately, there are not systematically a written blind test and an oral non-blind test for all subjects. In each track, we only consider the subjects for which there is both a compulsory written test and a compulsory oral test for all students[10]. This leaves us with a calibrated sample of 25,644 test scores (half written, half oral). Depending on the track, there are between three and six subjects for which all students have scores both at written and oral tests (see table 2). Note that some tests may be chosen as an option by students (see appendix tables A1 and A2). As a consequence, we cannot observe all the students in these tests. We have chosen to exclude these optional tests in our empirical analysis because, as these tests reveal students preferences, they may induce a strong selection of students who take them as well as particular grading practices by evaluators. Our results are nonetheless robust to

[8] Only a very small fraction refused to enter the ENS upon having been accepted.

[9] Students in preparatory schools are allowed to repeat their second year if they are not satisfied by the offers they got after taking the entrance exams of *Grandes Ecoles*.

[10] In rare cases, students take 2 written or oral tests in the same subject. In that case, we have averaged the candidates' scores over the two tests in order to keep only one observation per triplet *(student, subject, type)* where "type" distinguishes written from oral tests.

including these optional tests. The number of candidates that have taken both a non-optional written test and a non-optional oral test in each subject in each track is given in table 2. This number may vary slightly from a subject to another (within a track) because a few students did not present themselves to all tests (e.g. because of illness). Besides, the number of candidates is lower for tests on Latin/Ancient Greek and Foreign languages because we only kept data for students who chose the same language at both written and oral tests, so that both call for the same abilities[11].

Finally, scores at each written or oral tests in a given subject have been standardized to a distribution with zero mean and a unit standard deviation.

Indexes for subjects and tracks degree of feminization

We build an index I_s in order to characterize how "feminine" or "masculine" a given subject is. To keep the index simple, we consider the proportion of women among professors (*Professeurs des universités*) and assistant professors (*Maîtres de Conférences)* working in the corresponding field in French universities[12]. This choice is particularly relevant in our context because most of the students recruited by the ENS are going to become researchers. The value that takes our index for each subject is given in parenthesis in table 2, whose columns have been ordered according to this index[13].

We then build an index I_t that characterizes how "feminine" or "masculine" is a given track. To do so, we simply take a weighted average of our first subject-level index over all the subjects present in a given track, the weights being the actual coefficients that are applied to subjects when computing the student final averaged score and rank in the track. The value of this second index for each track is given in parenthesis in table 2, whose rows have been ordered according to this index. Here again, alternative indexes could be constructed, such as one corresponding to the share of female eligible candidates in each track. Taking this latter index rather than the former does not affect our results.

[11] 68% and 32% of the students in the Humanities track respectively chose Latin and Ancient Greek. Foreign languages are English (69%), German (24%), Spanish (4%) and other languages (3%).

[12] Statistics available at the French Ministry of Higher Education and Research website (http://media.enseignementsup-recherche.gouv.fr/file/statistiques/20/9/demog07fniv2_23520_49209.pdf). Keeping only professors or assistant professors to build our index does not affect our results.

[13] We have also tried to build a subjective index by averaging the perception of a sample of people around us that had scaled between 0 and 10 how they felt each subject was feminine. We finally discarded this index because of the difficulty to construct it from a random sample if individuals. However, non-surprisingly, results for both indexes were very similar, which shows that the proportion of female in academics in each field is a good measure of what people perceive as being a feminine or masculine subject or field.

We finally build a third index I_{st} giving the relative degree of feminization of a given subject in a given track by subtracting to the subject index the value the corresponding track index: $I_{st} = I_s - I_t$. The goal of this index is to capture the fact that for example chemistry is relatively feminine subject in the Physics-Chemistry track whereas it is a relatively masculine subject in the Biology track.

3. Empirical specifications

As candidates may share unobservable characteristics that are correlated to their gender and may affect their score, the gap between girls' and boys' average scores at oral examinations cannot be directly interpreted as a result of teachers' discrimination. In order to identify the role of teachers in students' grades, researchers usually implement difference-in-differences strategies. They compare for instance the score gap of the same candidates between two different subjects with different teachers (Dee 2007), or at the same subject between two different tests that are respectively blind and non-blind toward gender (Lindahl 2007; Lavy 2008). Implicitly, they assume students' individual effects to be fixed between both test's score so that their difference correctly identifies teachers' effects. Thus, the difference between boys' and girls' score gaps should give an unbiased estimate of teachers' gender discrimination.

Basic regressions (gender differences in oral-written score gap)

Similarly, we use the fact that our data on the ENS entrance exams contain both written anonymous tests and oral tests for a given eligible candidate in a given subject. The structure of the data with systematically one written and one oral test for each candidate in each subject makes it possible to use a difference-in-differences estimation strategy, similarly to Lavy (2008). More specifically, the score of candidate *i* in subject *j* is a function of gender (F), the oral nature of the test (O) and their interaction. Assuming a linear model, we can write:

$$S_{ijo} = \alpha_{ij} + \gamma_j O_{ijo} + \delta_j (F_i \times O_{ijo}) + \varepsilon_{ijo} \qquad (1)$$

where S_{ijo} is the score of candidate i at test of type o (written or oral) in subject j. F_i is an indicator equal to 1 for female candidates and O_{ijo} is an indicator equal to one for oral tests. α_{ij} is an individual fixed effect by subject that will take as value the score of candidate *i* at her written test in subject *j*. γ_j measures the difference between average scores at oral and written tests in subject j for men. δ_j is finally our parameter of interest: it measures the difference between oral and written tests in subject *j* for women, on top of the respective difference for men. As long as individual effects are assumed constant between written and oral tests, δ_j may be interpreted as the effect of the jury's bias toward girls in subject *j* (see Lavy, 2008, p. 2088 for details). This assumption does not hold for instance if girls are less competent than boys at oral exams (discussed later).

To simplify our empirical analysis and future exposition, we consider an equivalent of equation (1) in first differences. Noting $\Delta S_{ij} = S_{ijoral} - S_{ijwritten}$, we thus start by estimating:

$$\Delta S_{ij} = \gamma_j + \delta_j F_i + \varepsilon_{ij} \tag{2}$$

Since our data consists in a sample of 6 years pooled together, we have allowed γ_j to vary by year. However, since our goal is not to study across-time evolutions, we suppose that δ_j is constant over the period of observation (in order to maximize our statistical power). We can also add controls for individual characteristics that may be correlated to both gender and the first differences in scores (parents' occupations, age, former results at the *Baccalauréat* exam and former preparatory school):

$$\Delta S_{ij} = \gamma_j + \delta_j F_i + X_i + \varepsilon_{ij} \tag{3}$$

where X_i is a set of non-varying individual characteristics. Because these variables are only available since 2006, equation (2) may only be estimated on the 2006-2009 sub-sample.

The differences-in-differences nature of equation (2) means that account is taken of any ability measure that has a similar effect on the written and oral scores. However, if ability affects each of the scores differently, the difference of the two scores will depend on ability as well. We then get the following a model:

$$\Delta S_{ij} = \gamma_j + \delta_j F_i + \mu_j A_{ij} + \varepsilon_{ij} \tag{4}$$

where A_{ij} is the unobserved ability of candidate i in subject *j*. The fact that ability may be loaded differently in the written and oral test scores' equations will bias our results if the distribution of abilities is different for male and female candidates. However, we will provide some evidence showing that the distributions of scores at written tests are quite similar for male and female candidates. We will also try to control directly for ability in our empirical analysis by estimating equation (4) using as a proxy for A_{ij} the quartile of candidate i 's score in the score distribution of the written test in subject j.[14]

Within-candidate between-subjects differences

There may finally be an unobserved ability component A^{O}_{ij} that is specific to oral tests and that does not intervene in written tests. In that case, equation (2) would write:

$$\Delta S_{ij} = \gamma_j + \delta_j F_i + A^{O}_{ij} + \varepsilon_{ij} \quad (5)$$

We use the most interesting feature of our dataset – the fact that each candidate takes both a written and an oral test in several different subjects – to deal, at least partly, with this issue. To do this, we need to suppose that the oral ability component A^{O}_{ij} is common across subjects and can be written A^{O}_{i}. In that case, we can directly control for it by estimating a version of the first difference equation (2) with individual fixed effects:

$$\Delta S_{ij} = \gamma_j + \delta_j F_i + \alpha_i + \varepsilon_{ij} \quad (6)$$

where α_i capture the general ability of candidate i at oral tests. Identification now relies on variations of the oral/written score differences across subjects for female and male candidates. Due to this, equation (6) can only identify in each track the differential in scores between female and male candidates in all subjects but one taken as reference, and for which we impose the difference to be zero (see for example Dee, 2005 for a similar normalization). We finally estimate versions of equation (6) that include our controls for the general ability of the candidates (their quartile in the written test score distribution in each subject).

Our final exercise consists in nesting together all our estimates by track and subject using our two indexes for the feminine character of subjects and tracks. To do so, we estimate equations such as:

$$\Delta S_{ij} = \gamma_j + \delta(I_j \times F_i) + \alpha_i + \varepsilon_{ij} \quad (7)$$

where I_j is the index for the degree of feminization of subject j. In our empirical analysis, equivalents of equation (7) will also be estimated without controlling for students' general

[14] We try to avoid to control directly by the candidates' score at written tests on the left hand side because the variable would then appear on both side of the equation, rending our fixed-effect setting ineffective.

oral abilities α_i, as well as using our indexes for the degree of feminizations of tracks (I_t) and for the relative degree of feminization of a subject within a track (I_{st}).

4. Results

4.1 Gender differences between oral and written test scores by tracks

We start by applying our difference-in-differences estimation strategy (equation 2) at the track level. To do so, we simply average, in each track, the oral premium for girls in all subjects in which oral and written tests are both non optional (see table 2)[15]. We also estimate the oral premium for females at the level of the whole ENS entrance exam by pooling all tracks together. Our results show that the average difference between oral and written test scores at the ENS entrance exam for years 2004 to 2009 is significantly lower for girls (by about 5% of a standard deviation (s.d.) – see table 3, panel A, column 1). However, this differential varies strongly across tracks. Positive in the Math-Physics track (by about 10% of a s.d. – see column 2), the difference becomes negative in the Humanities track (by about 10% of a s.d. – see column 6). According to our index, the Math-Physics and Humanities track are respectively the most male-connoted and the most female-connoted tracks of the ENS entrance exam. It thus appears that discrimination, if any, goes in favor of girls in the most male-connoted tracks and in favor of boys in the most female-connoted tracks. Consistent with this theory, we do not find significant differences between female and male candidates oral premiums in the Physics-Chemistry, Biology-Geology and Social-Sciences tracks. These tracks indeed stand between Math-Physics and Humanities in terms of their degree of feminization.

The lower panel of table 4 gives the proportion of girls finally admitted in the ENS in each track during years 2004 to 2009, as well as the number of girls that would have been accepted if the exam had only consisted in the written exams. These statistics have been computed from candidates' rank at the exam, as well as from their rank at the eligibility step (i.e. after the written tests only). They allow us to confirm our regression results on the full sample of tests and to present quantified estimates of what might have been the consequences of

[15] All the following results are not only robust but strengthened by the inclusion of optional subjects such as Computer sciences in the Math-Physics track, or Geography in the Social Sciences and Humanities tracks.

discriminatory behaviors from the jury members on the final sex ratios in each track[16]. If the exam had stopped after the eligibility step, the proportion and number of girls among the admitted candidates would have been 4% higher (in relative terms) than the actual proportion and number of girls among the accepted candidates (panel B, column 1). However, this statistics varies again dramatically across tracks. In the Math-Physics track, the number of admitted girls is as high as 55% higher than what it would have been if the exam had stopped after the written tests. This number is still positive in the Physics-Chemistry track and gets negative in other tracks. Overall, results in panel B are consistent with our regression estimates presented in panel A.

One might worry that the distribution of abilities between girls and boys in the different tracks are so different that our gender comparisons are not relevant. Girls might for example be in the lower part of the ability distribution in the Math-Physics track whereas they are in the upper part of the ability distribution in the Humanities track. In that case, our results could simply reflect composition effects in the ability distribution combined with reversion to the mean or a variable return to ability along the ability distribution. Figure 1 gives the distribution of test scores for both males and females eligible candidates at written and oral tests in each track. When all tracks are considered together, the distributions of scores at written tests are remarkably similar for girls and boys[17] (see the first two graphs in figure 1). It is only at oral tests that the distribution of girls' test scores appears to be shifted leftward relative to the distribution of boys' test scores. The test scores distributions at written tests for males and females candidates are still very similar when we consider tracks separately. They are perfectly matched in the Physics-Chemistry and Humanities tracks whereas minor differences appear in other tracks. Finally, comparisons of the scores' distributions for boys and girls at oral and written tests confirm the pattern that emerged in table 3: in the Math-Physics track, the girls' distribution is shifted to the right at oral tests relative to that of boys whereas the opposite occurs in the Humanities track.

4.2 Gender differences between oral and written test scores by subject (and track)

---

[16] These ranks are computed by the exam board as a weighted average of all test scores in the exam, including optional tests and tests in subjects for which there is only a written or an oral test. Conversely, results presented on Table 3 panel A are estimated from non-weighted regression, giving an equal weight to each subject. However, weighting our regressions only strengthen our results since discrimination behaviors appear to be usually stronger in the most important subjects in each track (see table 4).

[17] The scores' distributions at written tests could also be computed on a larger sample that also includes candidates that were not eligible for oral tests. When doing so, we find that females are dominated by males in all tracks at written tests.

Girls seem to be favored in the most male-connoted tracks and boys in the most female-connoted tracks. We now investigate the existence of a similar pattern between the different subjects of a given track. Are girls favored in the most masculine subjects of a given track exam and boys in the most feminine subjects?

Our goal is twofold: comforting our results by track and solving some important identification issues that affect our results by track. Indeed, even if male and female candidates selected for the oral tests appear to have very similar abilities at written tests in all tracks (figure 1), they may be selected with respect to their oral abilities. The few women that have decided to major in Math-Physics despite strong social norms against such a choice may be for example especially self-confident or motivated, which in turn may affect their performance at oral tests more strongly than their performance at written tests. To tackle this problem, we now make comparisons within tracks, between the different subjects for which we have both non-optional written and oral test scores. Our estimates of the oral/written premium for girls relative to boys in the different subjects of a given track (see table 2) are obtained on the same sample of candidates. As a consequence, between subjects differences in the oral premium for girls in a given track cannot be attributed to unobservable gender differences in our sample of candidates.

Table 4, panel A, presents estimates obtained by regressing, in each track and subject, differences between oral and written test scores on an indicator equal to one for female candidates. For the sake of clarity, we have pooled together observations for all subjects in a given track and we have saturated the corresponding estimated equation with dummies for each subject in each year and dummies for each subject interacted with gender[18]. Both tracks and subjects are sorted according to their degree of feminization (according to our indexes). To give a better idea of the sense of our estimates, we have plotted them on a 3-d graph in figure 2.

Evidence globally supports the idea that within each track, girls are more favored in more male-connoted subjects. Perhaps not surprisingly, the premium for girls at oral tests in the Math-Physics track is almost entirely due to the Math subject in which females get an oral versus written test premium relative to males which is as high as 40% of a s.d. over the period 2004-2009. The oral versus written test premium is also positive, although non-significant, in

[18] We checked that our results are identical to what would be obtained by estimating one equation for each subject in each track.

the physics subject. Finally, this premium turns negative (although non-significant) in Foreign Languages which is, according to our index, almost the most feminine subject. If we move to other track, a similar pattern is observed. However, in more feminine tracks, estimates for all subjects are uniformly shifted downward and girls get discriminated. For example, the premium in the math subject does not appear in the Physics-Chemistry track and in the Social-Sciences track. In the Physics-Chemistry track, girls get a negative premium in chemistry which is the most feminine of the scientific subjects present in the track. A same pattern is observable in the Biology-Geology track where girls get a strong penalty (40% of a s.d.) in biology which is the most feminine scientific subject present in the track. In the Social-Sciences track, females seem to get a premium at oral tests relative to males in philosophy, which is the most masculine non-scientific subject of this (non-scientific) track. On the other hand, they are penalized in literature, which is conversely the most feminine subject in the track, even though the estimate remains non-significant. Female candidates experience a penalty in the oral foreign languages test relative to the written one in almost all tracks, and this penalty is particularly high in the Humanities track. In this track, girls are almost penalized everywhere and especially in the most feminine subjects.

It seems that a general pattern emerges from table 4, panel A. The premium for female candidates at oral tests decreases both when one moves rightward in a given row and downward in a given column. Of course, there are some exceptions. We think that most of the exceptions observed are due to the fact that comparisons are more relevant within scientific subjects in scientific tracks, and within non-scientific subjects in non-scientific tracks. The remaining exceptions may be due to some context specific elements, to the weakness of our indexes of feminization or to the lack of precision of some of our estimates. A case by case study would be required to understand exactly what happens in each test, which is certainly beyond the scope of this paper.

In order to directly control for candidates' general ability at oral tests, we have also estimated the effect of being a girl on the difference between oral and written test scores in specifications that include individual fixed effects (equation 6 – see table 4, panel B). The inclusion of these fixed effects implies that we can now only identify variations across subjects of premium for girls at oral tests relative to written tests. As a consequence, one subject has been chosen as a reference in each track. We have taken foreign languages as reference in all tracks[19] but the Social-Sciences track in which literature is the reference subject. The general pattern observed in panel A is still observable in this "triple-differences" framework. In the Math-Physics track, female candidates get a significantly higher oral versus written premium in math than in foreign languages. In the Physics-Chemistry (resp. Biology-Geology) track, there is still a penalty for female candidates in chemistry (resp. biology), which is the most female-connoted scientific subjects. Female candidates are favored in the most male-connoted non-scientific subject (philosophy) in the Social Sciences and Humanities tracks. Note that comparisons between tracks should not be made in table 4 panel B since the estimates only give within track variations between subjects.

We made several robustness checks for our results. Specifications without individual fixed effects (Panel A) are reproduced with individual controls for years 2006-2009 in table A4-Panel A and with controls for the candidates' initial ability in each subject (taken as the quartile of the written test scores distribution they belong to) in table A4-Panel B[20]. Specifications with fixed effects (Panel B) are reproduced with controls for the candidates' initial ability in each subject (taken as the quartile they belong to in the written test scores' distribution) in table A4-Panel C. Results in the 3 panels of table A4 are globally similar to those presented in table 4, showing that our results are not driven by gender differences in candidates' observable characteristics.

4.3 Gender differences between oral and written test scores depending on the degree of feminization of tracks and subjects

We finally nest together the disaggregated results presented in table 4, by estimating the effect of gender interacted linearly with our indexes of feminization (equation 7) on the full sample of ENS candidates. Consistent with our previous results, the oral premium for girls is significantly lower in the most feminine subjects (table 5, column 1 without fixed effects and column 6 with fixed effects). A 10 percentage points increase in the proportion of female scholars (both professors and assistant professors) in a field leads to a decrease of the oral

[19] We have done so to facilitate comparisons across tracks since foreign languages is the subject which appears in the largest number of tracks.

[20] We controlled for initial ability to check if our results were not only driven by mean reversion. If girls were better than boys at written tests on feminine subjects, the written-oral differential may be higher for girls on masculine subjects without any discrimination. Although controls added on table A4 and panel B may be sufficient to solve this problem, we also performed other analyses. First, we estimated gender gaps in written scores by subject and track. The results showed that there were significant gender gaps in anonymous written tests, but the estimates were not systematically correlated with our results (for instance, girls had higher grades at the written mathematic test of the Math-Physics track). Moreover, results presented in table 4 are robust to the inclusion of girls' average written test score in each subject as an additional control. Results available on request.

versus written premium for girls of about 7% of a s.d. in the corresponding subject. This is a strong effect: it means that the difference in oral premiums for girls between math, where 15% of professors are female, and foreign languages, where 56% of professors are female, is above 25% of a s.d. The oral premium for girls is also significantly lower in the most feminine tracks (column 2), with a 10% increasing in our track feminization index (which is an average of the track's subjects degree of feminization weighted by the coefficients of each subject in the exam) also leading to a 7% of a s.d. increase in the female oral/written premium. According to our linear specifications, the female oral/written premium would be around 20% in a hypothetical subject with no female scholars (first row, column 1), or in a hypothetical track where all subjects have no female scholars (first row, column 2).

When indexes of feminization for both tracks and subjects are included in the regression model, only the subject index remains significant (column 3). This indicates that the variations in the gender premium at oral tests are probably more driven by variations between subjects than by variations between tracks. However, when absolute degree of feminization of subjects is replaced by the relative degree of feminization of subjects within tracks (I_{st}), both this variable and the degree of feminization of tracks are significant determinants of the gender premium at oral tests (columns 5). This is an important result that summarizes well our analysis. It confirms that the premium for females at oral tests is affected by the degree of feminization of tracks, and that it is also affected on top of this first effect by the relative degree of feminization of each subject within the track.

Our results could still be driven by differences in students' abilities if female candidates turn to be better at oral tests with respect to written tests in more male-connoted subjects and/or tracks. The design of the natural experiment we use does not allow us to control directly for this potential bias. However, a recent literature has now established that negative stereotypes against a given social group affect this group performance negatively when its identity is revealed. In a famous experiment among Indian subjects that were assigned the task to solve mazes under economic incentives, Hoff and Pandey (2006) have shown that revealing the subjects' caste before the task was lowering the performance of the lower castes (e.g. the untouchables). Such behaviors have been observed in different contexts (e.g. Stone et al., 1999, concerning black students) and are likely to be explained by a decrease in self-confidence among subjects facing a stereotype threat (Cadinu et al. 2005). Directly related to our context, Spencer et al. (1999) have shown that, as compared to a benchmark situation, female performance is higher at difficult math tests when these tests are advertised as not producing gender differences (i.e. when the stereotype threat is lowered) and that it is lower when tests are advertised as producing gender differences (i.e. when the stereotype threat is increased). Overall, the literature strongly suggests that female performance at the ENS oral tests (where their type is revealed) as compared to written tests (where their type is not revealed) should be higher in the subjects and tracks in which the stereotype threat is the highest, i.e. the most male-connoted ones. In contrast, our results show the opposite. We thus conclude that if there are differences in oral abilities between subjects among the ENS candidates, these differences probably go against our results and lead us to underestimate the true discrimination made by the ENS jury.

5. Discussion

Our results show that professors discriminate in favor of the minority gender: girls are positively discriminated in majors and subjects identified as « masculine », while negatively discriminated in « feminine » tracks and subjects. This contributes to the literature by showing that the relative demand for students in science does not aggravate the existing gender gaps in the supply of students. However, our results do not show that the demand for females in science plays no role in the gender gap. Indeed, in our case, math professors discriminate in favor of girls, but they face a very segregated pool of candidates that contains only a few girls. Maybe would they discriminate against girls if they were more numerous among candidates. This study shows that the actual degree of segregation in the relative supply of females in science is larger than the "preferred gender gaps" on the demand side, not that the absolute demand is the same for female and male candidates. However, it seems clear that the reasons for the very large gender gaps across college majors may not be found exclusively on the demand side. Contrary to expectations if one draws straightforward interpretations from the literature on gender stereotypes, professors implement a strong positive discrimination in magnitude, even though not sufficient to compensate the huge gender gaps existing in the different majors. It raises many questions not only about the links between gender stereotypes and teacher grading behaviors, an issue that has been at the core of many scientific debates (Dee 2007; Lavy 2008), but also on the role played by professors in the gender gap (Carrell, Page & West 2010). We thus try to provide a couple of general explanations for our results. These explanations are likely to apply in a broad range of contexts and suggest that what we observe is not driven by some specificities of the institution in which takes place our natural experiment.

To begin with, the ENS and its jury members may implement a conscious affirmative action towards the minority gender in each major. In that case, our results would simply reflect that the ENS direction implements a policy towards gender equity and they would be arguably less interesting. However, the fact that we find very different estimates across subjects within a given track suggests that we observe more than an explicit policy in favor of the gender in minority in each track. Indeed, such a policy should probably lead to a similar premium for girls in all subjects of a given track. Another possibility is that the jury of the exam manipulates the candidates' scores *ex post* in order to increase (or decrease) the final number of admitted girls[21]. The easiest (and discrete) way to do so is to favor girls (or boys) in the subjects that have the highest coefficients in each track, which turn to be those in which we observe the largest oral versus written differentials between females and males candidates (see Tables 4, A1 and A2). However, if such strategic manipulations really occur, they should concern only the candidates that are close to the admission threshold. Indeed, the jury does not want to admit a candidate that is too far from the required level or reject a candidate that had performed very well. Based on this observation, we have tried to detect the existence of strategic manipulations at the admission threshold. The number of candidates accepted each year in each track is defined by law in advance[22]. This implies that the ENS entrance exam is in fact a contest. As a consequence, there is not any predefined admission threshold in terms of average score: only the rank matters. The score threshold is defined each year depending on the level of the candidates. We have computed it as the mean of the total scores of the first rejected and last admitted candidates in each track each year. We have then normalized the candidates' total scores in each track such that they have a unit standard deviation and such that the admission threshold correspond to a total score of 0 for all tracks and years. We first provide in figure 3 graphical evidence of possible discontinuities or changes in slope in the distribution of scores around the admission threshold. The admission threshold appears to be systematically located close to the mode of the total scores' distribution. However, the

[21] The idea of such an *ex post* manipulation of grades may appear awkward in the sense that it is against basic principles of equity. However, we know from our interviews that the ENS jury does such manipulations some years, but rarely and especially in the Math-Physics track. The justification they give for this is that when a normally non-admitted candidate was especially good in one particular subject and really impressed the examiner, the jury tries to push this candidate above the admission threshold if she is not too far and if the subject is important for this track. Of course, since the ENS entrance exam is actually a contest (the number of places is fixed), this means that another candidate will happen to be non-admitted.

[22] This is because the ENS is a public institution financed by the French government which, as a consequence, strictly supervises its functioning.

distributions do not present any clear sign of discontinuity at the admission threshold. To confirm this graphical diagnosis, we performed McCrary test (McCrary, 2008), as it is standard in the Regression Discontinuity Design (RDD) literature. In our context, McCrary test relies on two hypotheses. First, the distribution of the candidates' scores needs to be continuous in the absence of manipulation (this is a standard assumption in the RDD literature). Second, manipulation near the admission threshold needs to be "unilateral", in the sense that the ENS jury may increase the total score of some candidates to push them above the threshold, but will never decrease the total score of candidates in order to pull them below the threshold[23]. Under both hypotheses, manipulation can be detected by the presence of a discontinuity in the scores' distribution at the admission threshold. Even though the total scores' distribution appears to reach a peak and to be a bit irregular around the threshold, McCrary test did not detect a lack of continuity at the admission threshold for any track except for Math-Physics (see figure 4). The latter track may be the only one where some strategic discrimination occurs to improve the gender mix. Notice, however, that the small discontinuity detected at the admission threshold in this track is negative, which is counter-intuitive since we were expecting the jury to push some students above the threshold rather than the opposite. Despite this somehow puzzling exception, *ex post* strategic manipulation at the ENS entrance exam remains too limited to be detectable by standard analysis of the total scores' distributions[24].

In order to directly confirm that such strategic discrimination is not driving our results, we also checked that the jury bias toward the minority gender is not concentrated only on candidates who were close to the admission threshold at the end of the eligibility step. If our results were driven by strategic discrimination to improve gender mix, the jury would have chosen students at the middle of the underlying ability distribution and we should not find significant biases on the other students. However, when we divide our sample in three groups according to the candidates' ranks after the eligibility step, we also find the pattern exhibited in table 5 (i.e. that the gender gap in the written-oral differential varies with the tracks and

[23] Note that this second assumption was obviously verified in the original McCrary framework because manipulation at the threshold comes from the treated individuals themselves to move towards the preferred side of the threshold only. In our case, candidates can in principle be moved by the ENS jury in both directions. If the number of candidates moved by the ENS jury from under the threshold to above the threshold is equal to the number of candidates moved the other way around, the final scores' distribution under manipulation will still be continuous and manipulation will as a consequence be undetectable. However, our interviews with the ENS jury suggest that this second hypothesis is likely to be true: the jury does not feel comfortable with explicitly penalizing a candidate *ex post* whereas they may be willing to favor one in some cases.

[24] As a robustness check, we also performed McCrary tests for boys and girls separately, and we did not detect a lack of continuity at the admission threshold in any of these cases. Results available on request.

subjects degree of feminization) both for students located around and below the rank corresponding to the admission threshold (see table A6 reproducing columns 5 and 6 of table 5 on subsamples of the data). We thus conclude that the general pattern of increasing bias for girls with the track and subject's degree of masculinity cannot be explained by explicit affirmative action, that is, by a conscious policy of the ENS in favor of gender diversity.

We distinguish between three mechanisms that are likely to generate positive discrimination towards the minority gender. First, our results could be explained by a taste-based discrimination where professors have a preference for gender diversity. This preference may be due to the lack of girls or boys in their field. They may enjoy more interviewing a boy if they only work daily with girls. This mechanism is much more plausible than the conscious affirmative action policy explanation, as it is consistent with the differences between subjects that we find in each track, for example the fact that the same girl is negatively discriminated by biology professors (a field where girls are not underrepresented) while positively discriminated by geology professors. Furthermore, during the ENS entrance exams, the stake is not only to put a grade on an academic performance. Admitted students will enter into one of the top French higher education institution whose role is precisely to train students for top research careers. As a consequence, when they evaluate students' academic output, professors are simultaneously selecting people who are likely to become their peers within a few years. This situation differs strongly from examinations in contexts where candidates are not yet oriented towards a given career. In the case we study, a taste for gender diversity may have stronger effects on scores because professors directly affect the future gender mix in their field when they favor the gender in minority at the ENS entrance exams.

Professors could also give a premium to candidates of the minority gender if they care not only about the actual performance, but also about the gap between the actual and the expected performance with regard to gender stereotypes. If they expect girls to have lower abilities in mathematics than boys, professors may give them a premium for the same performance because it reveals their motivation and their ability to provide worthwhile efforts. Motivation and perseverance may indeed be valued by professors as they signal students' long-term potential. Such a mechanism seems likely at the ENS entrance exams where students are recruited for four years by researchers and are aimed to become researchers themselves, a profession that requires strong motivation and long-term investments (as the jury of the exam perfectly knows). Interestingly, this explanation is consistent with both dimensions of our results. To begin with, students from the minority gender may signal stronger motivation and self-investment by choosing a major for which they have lower academic abilities. Professors may thereby give a premium to girls (resp. boys) in the more "masculine" (resp. "feminine") tracks. Then, professors may not expect a given candidate in a given major to have the same abilities regarding its gender and the subject. Within each track, they may thus favor girls in the more "masculine" subject because the same performance level reveals higher abilities to get over difficulties. Our results are therefore also consistent with the idea that grades are partly based on the inferred student's perseverance. This second plausible mechanism is different from discrimination strictly speaking, i.e. professors favoring a less worthy group. Precisely, professors may not judge this group less worthy as the actual performance may not be the only criterion to define the "worthy" student: the expected long-term potential may also matter[25].

A last plausible mechanism worth mentioning is a specific kind of statistical discrimination (Phelps 1972; Arrow 1973) that can occur if the candidates' abilities are not perfectly observable during the ENS entrance exam tests. Arrow (1973) argues that discrimination can be rational even in the absence of both group-specific preferences and *ex-ante* differences in abilities between groups. As shown for the labor market with perfect information (Coate & Loury 1993; Moro & Norman 2004), this is because the beliefs of the employers concerning employees' abilities are going to be self-fulfilling: since the employees who are believed to be less able will be less rewarded *ex post*, their incentive to invest in human capital is lower and they will indeed be less able at equilibrium. The theory applies well in our context: if teachers and professors have stereotypes against girls in math, girls do not have strong incentives to invest in math (i.e. to enter a math major) and they finally happen to be (in average) less good than boys at math at equilibrium even though there were no initial differences in abilities between the two groups. But what about the few girls that overcome the initial adversity in math and try to major in math anyway? Conditional on being observed in a math majors, girls might actually be better than boys because they have already managed to jump the hurdle that stereotypes have raised in front of them. This mechanism is similar to the "belief flipping" described by (Fryer 2007) in the labor market: "If an employer discriminates against a group

[25] Previous French sociological research states that jury only reward pure talent at ENS entrance exam (Bourdieu & Passeron 1989). We do not oppose here the idea that professors are primarily looking for the highest talents. Nevertheless, according to ENS entrance exam jury members that we have interviewed, only a few students really stand out from the others and can be easily graded as excellent whereas the jury confessed that it is actually difficult to score the other more average candidates' performances at oral tests. The mechanism we describe concerns mainly these latter candidates, for whom other criteria such as the intrinsic motivation, perseverance and ability to provide future efforts may have an impact on scores.

of workers in her initial hiring, she may actually favor the successful members of that group […]". Fryer's model can easily be applied to our setting: professors may have negative stereotypes against the general population of girls with regard to their abilities in math, but a positive prior towards the 9% of women who were successful enough to be eligible for oral tests at the ENS entrance exams in the Math-Physics major. If the candidates' abilities are not perfectly observable during the ENS entrance exams, gender can be rationally used at oral tests as an additional piece of information concerning these abilities. In that case, our results could reflect pure statistical discrimination, but after a belief-flipping (*à la* Fryer) occurred, that is in a context where the minority gender is believed to be better because it has faced a stronger initial selection.

As a conclusion, our results exhibit grade premiums toward the minority gender. This could be plausibly explained by three mechanisms: an unconscious taste-based discrimination with preference for diversity, a reward for high perseverance and motivation, or a statistical discrimination after a belief-flipping occurred. We are not able to disentangle these explanations using our data and this sounds a promising area for future research. However, our paper contributes to the literature on gender discrimination as it underlines the complexity of the relationships between stereotypes and discrimination, as well as the role of professors in the gender gap.

Bibliography

Arrow, KJ 1973, 'The theory of discrimination', in *Discrimination in Labor Markets*, Princeton University Press, Princeton.

Bernard, ME 1979, 'Does Sex Role Behavior Influence the Way Teachers Evaluate Students?', *Journal of Educational Psychology*, 1979, pp. 553-562.

Bettinger, EP & Long, BT 2005, 'Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students.', *The American Economic Review*, 2005, pp. 152-157.

Blank, R 1991, 'The effects of double-blind versus single-blind refereeing: experimental evidence from the American economic review', *The American Economic Review*, 1991, pp. 1041-67.

Bourdieu, P & Passeron, J-C 1989, *La noblesse d'Etat: grandes écoles et esprit de corps*, Les Editions de Minuit, Paris.

Brown, C & Cororan, M 1997, 'Sex-Based Differences in School Content and the Male-Female Wage Gap.', *Journal of Labor Economics*, 1997, pp. 431-465.

Cadinu, M, Maass, A, Rosabianca, A & Kiesner, J 2005, 'Why Do Women Underperform under Stereotype Threat?', *Psychological Science*, 2005, pp. 572-578.

Canes, BJ & Rosen, HS 1995, 'Following in Her Footsteps? Women's Choices of College Majors and Faculty Gender Composition', *Industrial and Labor Relations Review*, 1995, pp. 486-504.

Carrell, SE, Page, ME & West, JE 2010, 'Sex and Science: How Professor Gender Perpetuates The Gender Gap', *The Quarterly Journal of Economics*, 2010, pp. 1101-1144.

Carr, M, Jessup, DL & Fuller, D 1999, 'Gender Differences in First-Grade Mathematics Strategy Use: Parent and Teacher Contributions', *Journal for Research in Mathematics Education*, 1999, pp. 20-46.

Coate, S & Loury, GC 1993, 'Will Affirmative-Action Policies Eliminate Negative Stereotypes?', *The American Economic Review*, 1993, pp. 1220-40.

Croson, R & Gneezy, U 2009, 'Gender Differences in Preferences', *Journal of Economic Literature*, 2009, pp. 448-474.

Deaux, K & LaFrance, M 1998, 'Gender', in *The handbook of social psychology*, McGraw-Hill, New York.

Dee, TS 2007, 'Teachers and the Gender Gaps in Student Achievement', *Journal of Human Resources*, 2007.

Dusek, JB & Joseph, G 1983, 'The bases of teacher expectancies: A meta-analysis', *Journal of Educational Psychology*, 1983, pp. 327-346.

Fryer, RG 2007, 'Belief Flipping in a Dynamic Model of Statistical Discrimination', *Journal of Public Economics*, vol 91, no. 5-6, pp. 1151-1166.

Gardecki, R & Neumark, D 1998, 'Women Helping Women? Role Model and Mentoring Effects on Female Ph.D. Students in Economics', *Journal of Human Resources*, 1998, pp. 220-246.

Glick, P, Wilk, K & Perreault, M 1995, 'Images of Occupations: Components of Gender and Status in Occupational Stereotypes', *Sex Roles*, 1995.

Goldin, C & Rouse, C 2000, 'Orchestrating impartiality: the impact of ‘blind' auditions on female musicians.', *The American Economic Review*, 2000, pp. 715-742.

Hoffman, F & Oreopoulos, P 2009, 'A Professor Like Me: The Influence of Instructor Gender on College Achievement', *Journal of Human Resources*, 2009.

Hoff, K & Pandey, P 2006, 'Discrimination, Social Identity and Durable Inequalities', *American Economic Review*, 2006, pp. 206-211.

Lavy, V 2008, 'Do gender stereotypes reduce girls' or boys' human capital outcomes?', *Journal of Public Economics*, vol 92, pp. 2083-2105.

Lindahl, E 2007, 'Does gender and ethnic background matter when teachers set school grades? Evidence from Sweden', Working Paper, Uppsala University.

Madon, S, Jussim, L, Keiper, S, Eccles, J, Smith, A & Palumbo, P 1998, 'The accuracy and power of sex, social class, and ethnic stereotypes: a naturalistic study in person perception', *Personality and Social Psychology Bulletin*, 1998, pp. 1304-1318.

McCrary, J 2008, 'Manipulation of the running variable in the regression discontinuity design: A density test', *Journal of Econometrics*, 2008, pp. 698-714.

Moro, A & Norman, P 2004, 'A general equilibrium model of statistical discrimination', *Journal of Economic Theory*, 2004, pp. 1-30.

National Science Foundation 2006, 'Science and Engineering Degrees: 1966–2004', Manuscript, Division of Science Resources Statistics, National Science Foundation, NSF 07-307.

Phelps, ES 1972, 'The Statistical theory of Racism and Sexism', *The American Economic Review*, 1972, pp. 659-61.

Rothstein, D 1999, 'Do Female Faculty Influence Female Students Educational and Labor Market Attaintments?', *Industrial and Labor Relations Review*, 1999, pp. 185-194.

Rowsey, RE 1997, 'The Effects of Teachers and Schooling on the Vocational Choice of University Research Scientist', *School Science and Mathematics*, 1997, pp. 20-26.

Spencer, SJ, Steele, CM & Quinn, DM 1999, 'Stereotype Threat and Women's Math Performance', *Journal of Experimental Social Psychology*, 1999, pp. 4-28.

Stone, J, Lynch, CI, Sjomeling, M & Darley, JM 1999, 'Stereotype Threat Effects on Black and White Athletic Performance', *Journal of Personality and Social Psychology*, 1999, pp. 1213-27.

Tiedemann, J 2000, 'Parents' gender stereotypes and teachers' beliefs as predictors of children' concept of their mathematical ability in elementary school', *Journal of Educational Psychology*, 2000, pp. 144-151.

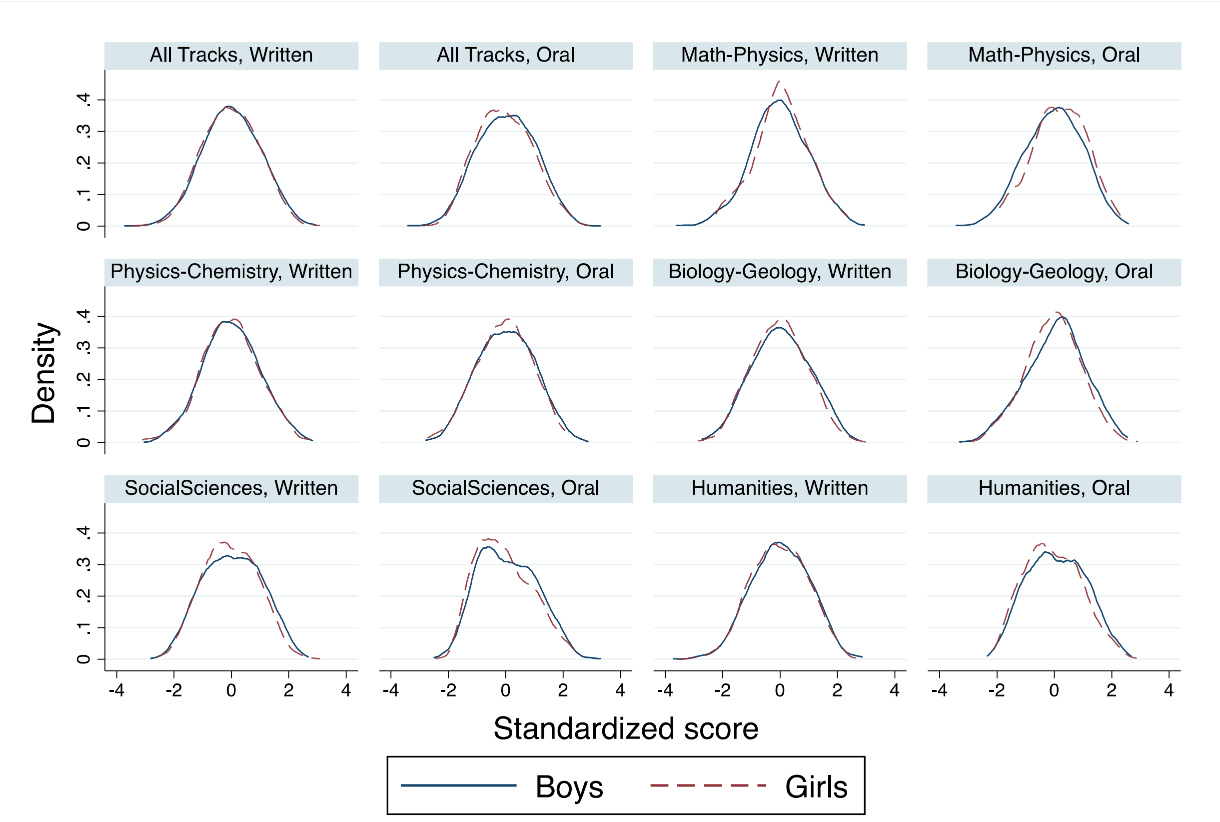
Weinberger, CJ 1998, 'Race and Gender Wage Gaps in the Market for Recent College Graduates.', *Industrial Relations*, 1998, pp. 67-84.

Weinberger, CJ 1999, 'Mathematical College Majors and the Gender Gap in Wages', *Industrial Relations*, 1999, pp. 407-13.

Weinberger, CJ 2001, 'Is Teaching More Girls More Math the Key to Higher Wages?', in *Squaring Up: Policy Strategies to Raise Women's Incomes in the U.S.*, University of Michigan Press.

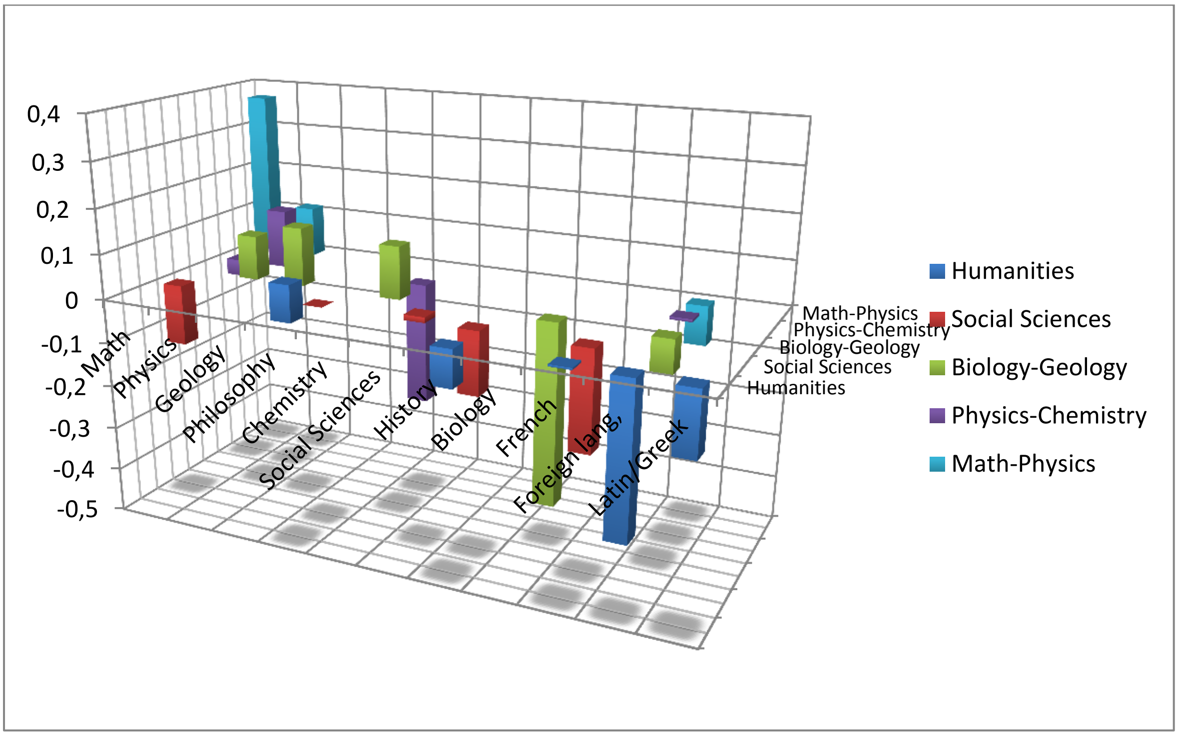
Tables and Figures

Figure 1: Kernel density estimates of scores at written and oral tests, by track



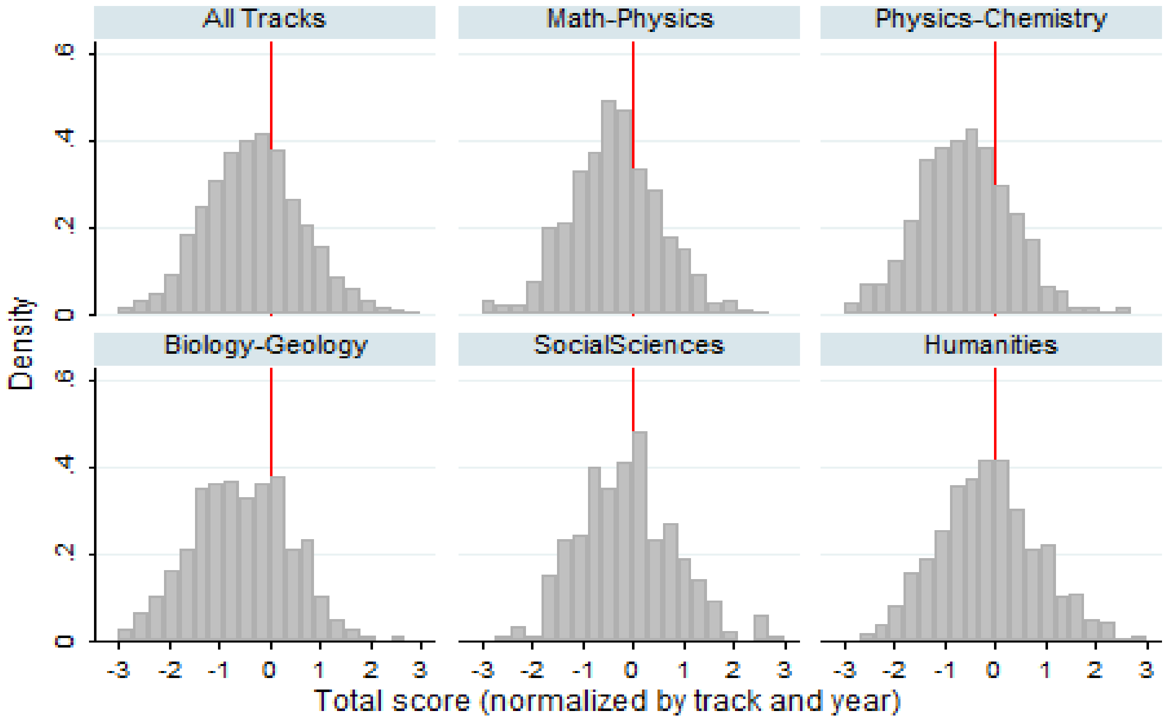
Notes: Kernel density estimates using Epanechnikov kernel function on Stata 12.0 software. The half-width of the kernel is an "optimal" width calculated automatically by the software, i.e. the width that would minimize the mean integrated squared error if the data were Gaussian and a Gaussian kernel was used.

Figure 2: The oral versus written premium for female in each track (graphical representation of the estimates of Table 4-Panel A).



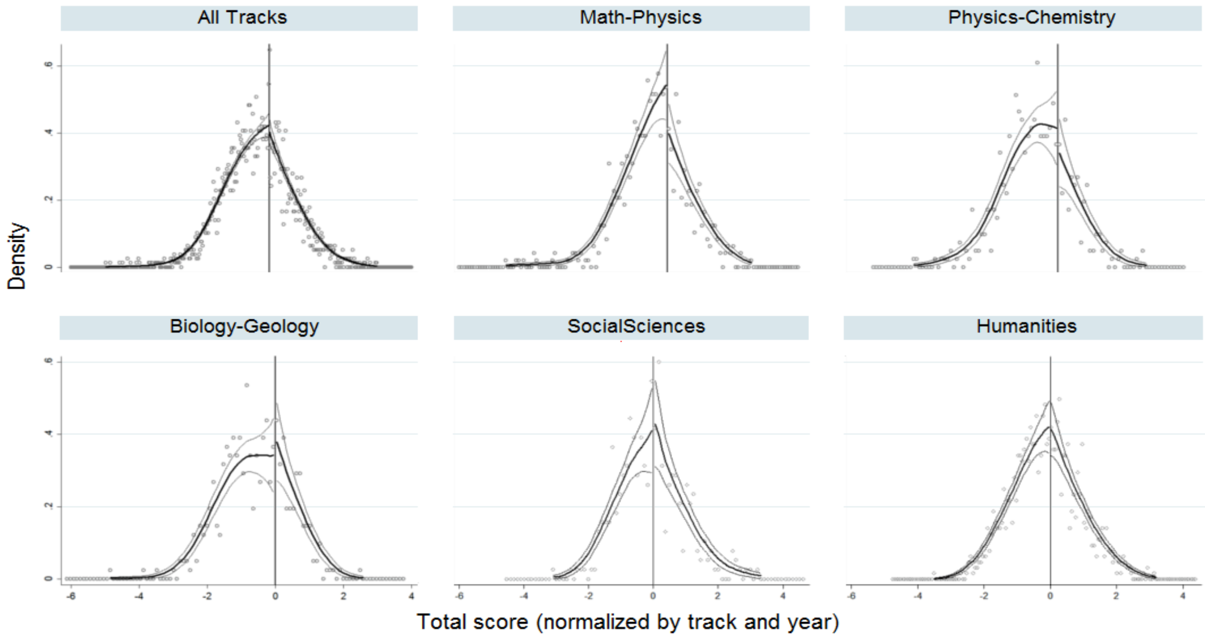
Note: Subjects are reported on the x-axis and tracks are reported on the y-axis. Subjects and tracks have been ordered according to our feminization indexes. Estimates presented on Table 4 – panel A are reported On the z-axis.

Figure 3: Distribution of students' total scores in each track



Note: The distributions of the candidates' total scores have been normalized in each track for each year (2004-2009) such that (i) the admission threshold always corresponds to a score of 0 (vertical bar), (ii) they have a standard deviation equal to 1.

Figure 4: McCrary test of a discontinuity at the admission threshold in each track



Note: The distributions of the candidates' total scores have been normalized in each track for each year (2004-2009) such that (i) the admission threshold always corresponds to a score of 0 (vertical bar), (ii) they have a standard deviation equal to 1.The McCrary works as follows: (i) smooth the total scores' distribution below and above the admission threshold, (ii) compute the confidence interval of the smoothed distributions, (ii) test if there is a significant discontinuity in the total scores' distribution at the admission threshold. See McCrary (2007) for details.

Table 1 : Descriptive statistics - Eligible candidates by track (2004-2009)

Track	All tracks	Math-Physics	Physics-Chemistry	Biology-Geology	Social Sciences	Huma-nities
Total eligible candidates	3068	747	506	438	335	1042
Average per year	511	125	84	73	56	174
Average admitted per year	184	42	21	21	25	75
% Girls among eligible candidates	40%	9%	16%	56%	53%	64%
% Admitted among eligible candidates	36%	34%	25%	29%	45%	43%
% Girls among admitted candidates	40%	12%	13%	44%	47%	59%

Table 2: Description of the subjects for which both a written and an oral test are available, by exam track

Track / Subject	Math-Physics (0.216)	Physics-Chemistry (0.269)	Biology-Geology (0.342)	Social Sciences (0.362)	Humanities (0.435)
Math (0.152)	**1480**	**956**	Written	**670**	
Computer Sciences (0.192)	Option				
Physics (0.213)	**1474**	**982**	**836**		
Geology (0.250)			**828**		
Philosophy (0.257)				**668**	**2070**
Geography (0.319)				Option	Option
Chemistry (0.331)		**978**	**836**		
Social Sciences (0.335)				**666**	
History (0.389)				**666**	**2070**
Biology (0.432)			**830**		
Literature (0.535)				**666**	**2073**
Latin/Ancient Greek (0.547)				Option	**1786**
Foreign languages (0.565)	**1452**	**958**	**832**	**333**	**1878**

Note: Size samples are given for the subject that we keep in our empirical analysis. "Written" means that there is only a written test for the subject. "Option" means that the subject is optional at the written test, oral test or at both. A blank is left in the corresponding box when a subject does not belong to a given track exam. Data for Latin/Ancient Greek and Foreign languages are only kept for students who chose the same language at written and oral tests. 68% and 32% of Humanities students respectively chooses Latin and Ancient Greek. Foreign languages are English (69%), German (24%), Spanish (4%) and other languages (3%). Indexes of feminization are given in parenthesis for each subject and each track. Subjects and tracks are ordered according to these indexes.

Table 3: Gender bias in oral tests by track

Panel A: Gender and differences between oral and written test scores- by track (2004-2009)

Track	all	Math-Physics	Physics-Chemistry	Biology-Geology	Social Sciences	Humanities
	(1)	(2)	(3)	(4)	(5)	(6)
Girl	-0.051**	0.131*	-0.045	-0.028	-0.047	-0.092**
	(0.024)	(0.079)	(0.070)	(0.054)	(0.061)	(0.036)
Controls	year*subject*track	year*subject	year*subject	year*subject	year*subject	year*subject
Observations	12,822	2,198	1,937	2,081	1,668	4,938
R-squared	0.000	0.001	0.000	0.000	0.000	0.001

Panel B: Proportion of female among accepted candidates considering oral and/or written tests

	all	Math-Physics	Physics-Chemistry	Biology-Geology	Social Sciences	Humanities
N admitted girls (a)	438	29	17	56	71	265
% among all admitted candidates	39.60%	11.60%	13.49%	44.44%	47.02%	58.50%
Counterfactual N admitted girls just after the eligibility step (b)	458	18	15	62	77	286
% among all counterfactual admitted students	41.41%	7.50%	11.90%	49.21%	49.04%	61.11%
Relative variation between (a) and (b)	*-4%*	*55%*	*13%*	*-10%*	*-4%*	*-4%*

Note: Panel A - The dependent variable is the candidates' difference between the oral and written test scores in each subject in which written and an oral tests are both non-optional. The number of observations is thus for each track the number of candidates times the number of subjects. Robust Standard errors in parentheses.
Panel B – The counterfactual is the number of girls who would have been admitted if the exam was only made up by the eligibility step (anonymous written tests only). It is based on the eligibility rank computed by the exam board to determine the pool of eligible students, to which we applied the final admission threshold of each track. We estimated then the number of girls within the resulting counterfactual pool of admitted students.
**** p<0.01, ** p<0.05, * p<0.1*

Table 4 : Gender differences between oral and written test scores - by subject and track (2004-2009)

Panel A : No controls

Track	Math-Physics (0.216)	Physics-Chemistry (0.269)	Biology-Geology (0.342)	Social Sciences (0.362)	Humanities (0.435)
	(1)	(2)	(3)	(4)	(5)
Math (0.152)	0.369***	-0.037		-0.137	
	(0.115)	(0.155)		(0.091)	
Physics (0.213)	0.113	0.131	0.099		
	(0.169)	(0.147)	(0.124)		
Geology (0.250)			0.131		
			(0.121)		
Philosophy (0.257)				0.253*	0.081
				(0.150)	(0.080)
Chemistry (0.331)		-0.278**	0.118		
		(0.141)	(0.121)		
Social Sciences (0.335)				0.012	
				(0.144)	
History (0.389)				-0.141	-0.083
				(0.142)	(0.078)
Biology (0.432)			-0.417***		
			(0.137)		
Literature (0.535)				-0.224	-0.004
				(0.149)	(0.088)
Latin/Ancient Greek (0.547)					-0.140*
					(0.072)
Foreign languages (0.565)	-0.089	0.006	-0.074		-0.339***
	(0.117)	(0.112)	(0.089)		(0.082)
Observations	2,198	1,937	2,081	1,668	4,938
R-squared	0.004	0.003	0.008	0.005	0.004
Year*subject dummies	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	No	No	No	No

*Note: The dependant variable is the candidate's difference between the oral and written test scores. Interactions between the girl dummy and each subject dummies are estimated and reported on the table. Indexes of feminization are given in parenthesis for each subject and each track. Subjects and tracks are ordered according to these indexes. Robust Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 4 : Gender differences between oral and written test scores - by subject and track (2004-2009)

Panel B: Controls for individual fixed effects

Track	Math-Physics (0.216)	Physics-Chemistry (0.269)	Biology-Geology (0.342)	Social Sciences (0.362)	Humanities (0.435)
	(1)	(2)	(3)	(4)	(5)
Math (0.152)	0.453**	-0.038		0.079	
	(0.178)	(0.199)		(0.181)	
Physics (0.213)	0.199	0.113	0.171		
	(0.200)	(0.190)	(0.156)		
Geology (0.250)			0.199		
			(0.143)		
Philosophy (0.257)				0.468**	0.430***
				(0.201)	(0.113)
Chemistry (0.331)		-0.283	0.192		
		(0.186)	(0.152)		
Social Sciences (0.335)				0.234	
				(0.197)	
History (0.389)				0.082	0.269**
				(0.199)	(0.112)
Biology (0.432)			-0.335**		
			(0.155)		
Literature (0.535)				REFERENCE	0.347***
					(0.118)
Latin/Ancient Greek (0.547)					0.197*
					(0.113)
Foreign languages (0.565)	REFERENCE	REFERENCE	REFERENCE		REFERENCE
Observations	2,198	1,937	2,081	1,668	4,938
R-squared	0.361	0.273	0.251	0.225	0.213
Year*subject dummies	Yes	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable is the candidate's difference between the oral and written test scores. Interactions between the girl dummy and each subject dummies are estimated with individual fixed effects (literature is the reference subject for the Social Sciences track ; foreign language for all other tracks) and reported on the table. Indexes of feminization are given in parenthesis for each subject and each track. Subjects and tracks are ordered according to these indexes. Robust standard errors in parentheses.

**** p<0.01, ** p<0.05, * p<0.1*

Table 5 : Estimated Gender Bias with indexes for Subjects and Tracks Degree of Feminization

	(1)	(2)	(3)	(4)	(5)	(6)
Girl	0.225***	0.208*	0.274**	-0.039	0.274**	
	(0.066)	(0.123)	(0.125)	(0.024)	(0.125)	
GirlI_j*	-0.707***		-0.678***			-0.603***
	(0.158)		(0.165)			(0.158)
Girl*I_t		-0.698**	-0.164		-0.841**	
		(0.330)	(0.343)		(0.334)	
Girl*I_{jt}				-0.631***	-0.678***	
				(0.163)	(0.165)	
Observations	12,822	12,822	12,822	12,822	12,822	12,822
R-squared	0.002	0.001	0.002	0.001	0.002	0.247
Track	all	all	all	all	all	all
Individual fixed effects	No	No	No	No	No	Yes
year*subject controls	Yes	Yes	Yes	Yes	Yes	Yes

*Note: The dependent variable is the candidate's difference between the oral and written test scores. "Is" is the subject feminization index, "It" the track feminization index and "I_{st}" their difference. Robust Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Appendix Tables

Table A1 : Description of the settings of ENS entrance exam in scientific tracks

Track	Math-Physics		Physics-Chemistry		Biology-Geology	
Speciality	Math-Physics	Computer Sciences	Physics	Chemistry	Biology	Geology
Written tests for all candidates	Math 1 (6)	Math 1 (6)	Physics (6)	Physics (6)	Biology (7)	Biology (4)
	Physics (6)	Physics (5)	Chemistry (6)	Chemistry (6)	Chemistry (4)	Chemistry (3)
	Math 2 (4)	Computer Sciences (5)	Math (5)	Math (5)	Physics (2)	Physics (3)
					Geology (2)	Geology (5)
Written tests for eligible candidates only	French (8)	French (8)	French (8)	French (8)	French (8)	French (8)
	FL 1 (3)	FL 1 (3)	FL 1 (3)	FL 1 (3)	FL 1 (3)	FL 1 (3)
	FL 2 (3)	FL 2 (3)	FL 2 (3)	FL 2 (3)	FL 2 (3)	FL 2 (3)
					Math (16)	Math (16)
Oral tests for eligible candidates only	Math 1 (25)	Math 1 (20)	Physics 1 (20)	Physics 1 (24)	Biology (25)	Biology (17)
	Math 2 (15)	Math 2 (10)	Chemistry 1 (20)	Chemistry 1 (20)	Geology (12)	Geology (20)
	Physics 1 (10)	Physics 1 (20)	Physics 2 (8)	Chemistry 2 (8)	Physics (16)	Physics (16)
	Physics 2 (20)	Computer Sciences (20)	Math (20)	Math (16)	Chemistry (16)	Chemistry (16)
			Physics lab work (12)	Physics lab work (12)	Biology or Chemistry lab work (12)	Biology or Chemistry lab work (12)
			Chemistry lab work (8)	Chemistry lab work (8)		
	SPW (8)	SPW (8)	SPW (8)	SPW (8)	SPW (15)	SPW (15)
	FL (3)	FL (3)	FL (3)	FL (3)	FL (3)	FL (3)

Note: Tests' weights in parenthesis.. Tests kept in the final sample are underlined.
FL = Foreign Language. SPW = Supervised Personal Work ("TIPE")

Table A2 : Description of the settings of ENS entrance exam in Social sciences and Humanities

Track	Social Sciences	Humanities
Written tests for all candidates	History (3) Philosophy (3) Literature (3) Social Sciences (3) Maths (3) Specialty subject[1] (3)	History (3) Philosophy (3) Literature (3) Foreign language (3) Latin/Ancient Greek (3) Specialty subject[2] (3)
Oral tests for eligible candidates only	History (2)[3] Philosophy (2)[3] Literature (2)[3] Foreign language (2)[3] Social Sciences (2)[3] Maths (2)[3] Specialty subject[1] (3)	History (2)[3] Philosophy (2)[3] Literature (2)[3] Foreign language (2)[3] Latin/Ancient Greek (2)[3] Specialty subject[2] (3)

Note: Tests' weights in parenthesis.

1 : The Specialty subjects chosen by candidates from the Social Sciences track should be drawn from the following list : Latin, Ancient Greek, Foreign Language, Geography. For the oral test, Social Sciences may also be chosen by eligible candidates. Eligible candidates may choose a different Specialty subject for the written and oral tests.

2 : The Specialty subjects chosen by candidates from the Humanities track : Latin, Ancient Greek, Literature, Philosophy, Music studies, Art studies, Theater studies, Film studies, Foreign Language, Geography. Eligible candidates may choose a different Specialty subject for the written and oral tests.

3 : Eligible candidates from the Social Sciences track (resp. Humanities track) choose one of these 6 (resp. 5) subject to be weighted by 3 instead of 2.

Table A3: Observable characteristics of eligible female and male candidates (2006-2009 only)

Track	Math-Physics			Physics-Chemistry			Biology-Geology			Social Sciences			Humanities		
	Boys	**Girls**	**Diff**	**Boys**	**Girls**	**Diff**	**Boys**	**Girls**	**Diff**	**Boys**	**Girls**	**Diff**	**Boys**	**Girls**	**Diff**
Low or middle social background	19%	10%		28%	22%		37%	30%		23%	16%		29%	22%	**
High Honors *Baccalaureat* graduate	68%	93%	***	60%	71%		63%	82%	***	73%	74%		69%	77%	**
"High quality" preparatory school	72%	72%		53%	59%		58%	56%		87%	85%		88%	89%	
Repeater at preparatory cursus	38%	34%		42%	54%	*	20%	15%		50%	51%		57%	63%	
N	453	44		278	59		133	171		107	117		236	456	

*Note - The "Low social background" dummy equals 1 if the candidate's father belongs to the middle or lower class regarding its occupation. The "Highest Honours Baccalaureat graduate" dummy equals 1 if the candidate graduated the French Baccalaureat exam at the end of high school with a grade superior or equals to 16 over 20. The "High quality preparatory school" equals 1 if the candidate comes from a preparatory school where at least 4 students managed to be admitted to the ENS during the 2006-2009 period, i.e 1 student per year in the average. The "Repeater at preparatory cursus" equals 1 if the candidate has repeated its second preparatory year to resit the "Grandes Ecoles" entrance exams. For each variable and track, the gender gap is tested by Pearson's chi-square test and the significance level is reported on the "Diff" column. *** : Significant at 1%. ** : Significant at 5%. * : Significant at 10%*

Table A4 : Gender differences between oral and written test scores - by subject and track (2004-2009)

Panel A : Controls for individual characteristics (2006-2009 samples only)

Track	Math-Physics (0.216)	Physics-Chemistry (0.269)	Biology-Geology (0.342)	Social Sciences (0.362)	Humanities (0.435)
	(1)	(2)	(3)	(4)	(5)
Math (0.152)	0.399***	-0.175		-0.121	
	(0.151)	(0.192)		(0.116)	
Physics (0.213)	-0.150	0.158	0.131		
	(0.217)	(0.168)	(0.147)		
Geology (0.250)			0.265*		
			(0.144)		
Philosophy (0.257)				0.212	0.132
				(0.181)	(0.102)
Chemistry (0.331)		-0.336**	0.091		
		(0.147)	(0.144)		
Social Sciences (0.335)				-0.122	
				(0.180)	
History (0.389)				-0.157	-0.033
				(0.174)	(0.099)
Biology (0.432)			-0.328**		
			(0.166)		
Literature (0.535)				-0.272	0.078
				(0.182)	(0.110)
Latin/Ancient Greek (0.547)					-0.091
					(0.092)
Foreign languages (0.565)	-0.114	0.016	-0.009		-0.419***
	(0.124)	(0.140)	(0.109)		(0.105)
Observations	1,402	1,266	1,423	1,108	3,237
R-squared	0.018	0.026	0.021	0.020	0.015
Year*subject dummies	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes

*Note: The dependant variable is the candidate's difference between the oral and written test scores. Interactions between the girl dummy and each subject dummies are estimated and reported on the table. Individual characteristics controls are 6 father's and 6 mother's occupation dummies, a dummy for repeater students at preparatory cursus, 4 dummies for "Baccalaureat" distinction levels, and a dummy for "High quality" preparatory school. Robust Standard errors in parentheses. Indexes of feminization are given in parenthesis for each subject and each track. Subjects and tracks are ordered according to these indexes. *** p<0.01, ** p<0.05, * p<0.1*

Table A4 : Gender differences between oral and written test scores - by subject and track (2004-2009)

Panel B : Controls for initial ability

Track	Math-Physics (0.216)	Physics-Chemistry (0.269)	Biology-Geology (0.342)	Social Sciences (0.362)	Humanities (0.435)
	(1)	(2)	(3)	(4)	(5)
Math (0.152)	0.376***	-0.088		-0.222**	
	(0.101)	(0.133)		(0.095)	
Physics (0.213)	0.108	-0.091	0.025		
	(0.149)	(0.131)	(0.110)		
Geology (0.250)			0.045		
			(0.108)		
Philosophy (0.257)				0.158	-0.051
				(0.128)	(0.072)
Chemistry (0.331)		-0.226*	0.061		
		(0.122)	(0.108)		
Social Sciences (0.335)				-0.002	
				(0.126)	
History (0.389)				-0.071	-0.178**
				(0.123)	(0.070)
Biology (0.432)			-0.327***		
			(0.117)		
Literature (0.535)				-0.167	0.001
				(0.126)	(0.077)
Latin/Ancient Greek (0.547)	0.046	0.153	0.049		-0.229***
	(0.103)	(0.111)	(0.085)		(0.072)
Foreign languages (0.565)					-0.110
					(0.067)
Observations	2,198	1,937	2,081	1,668	4,938
R-squared	0.234	0.287	0.332	0.349	0.322
Year*subject dummies	Yes	Yes	Yes	Yes	Yes
Controls for initial ability	Yes	Yes	Yes	Yes	Yes

*Note: The dependent variable is the candidate's difference between the oral and written test scores. Interactions between the girl dummy and each subject dummies are estimated and reported on the table. The initial ability control is the written score's quartile. Indexes of feminization are given in parenthesis for each subject and each track. Subjects and tracks are ordered according to these indexes. Robust Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table A4 : Gender differences between oral and written test scores - by subject and track (2004-2009)

Panel C : Controls for individual fixed effects and initial ability

Track	Math-Physics (0.216)	Physics-Chemistry (0.269)	Biology-Geology (0.342)	Social Sciences (0.362)	Humanities (0.435)
	(1)	(2)	(3)	(4)	(5)
Math (0.152)	0.310**	-0.244		-0.072	
	(0.152)	(0.178)		(0.158)	
Physics (0.213)	0.043	-0.265	-0.035		
	(0.169)	(0.176)	(0.139)		
Geology (0.250)			-0.025		
			(0.130)		
Philosophy (0.257)				0.311*	0.167*
				(0.168)	(0.099)
Chemistry (0.331)		-0.379**	0.003		
		(0.167)	(0.134)		
Social Sciences (0.335)				0.163	
				(0.166)	
History (0.389)				0.095	0.054
				(0.167)	(0.097)
Biology (0.432)			-0.369***		
			(0.136)		
Literature (0.535)					0.231**
					(0.101)
Latin/Ancient Greek (0.547)					0.121
					(0.100)
Foreign languages (0.565)	REFERENCE	REFERENCE	REFERENCE		REFERENCE
Observations	2,198	1,937	2,081	1,668	4,938
R-squared	0.553	0.510	0.560	0.541	0.509
Fixed effects	Individual	Individual	Individual	Individual	Individual
Controls	Initial ability	Initial ability	Initial ability	Initial ability	Initial ability

*Note: The dependent variable is the candidate's difference between the oral and written test scores. Interactions between the girl dummy and each subject dummies are estimated with individual fixed effects (foreign language is the reference subject) and reported on the table. The initial ability control is the written score's quartile. Indexes of feminization are given in parenthesis for each subject and each track. Subjects and tracks are ordered according to these indexes. Robust Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table A5: Estimated Gender Bias with indexes for Subjects Degree of Feminization by track

	(1)	(2)	(3)	(4)	(5)	(6)
Girl	0.225***		0.418***		-0.011	
	(0.066)		(0.153)		(0.160)	
Girl*Is	-0.707***	-0.603***	-0.925**	-0.915**	-0.105	-0.099
	(0.158)	(0.158)	(0.388)	(0.414)	(0.408)	(0.432)
Observations	12,822	12,822	2,198	2,198	1,937	1,937
R-squared	0.002	0.247	0.003	0.360	0.000	0.271
Track	all	all	Math-Physics	Math-Physics	Physics-Chemistry	Physics-Chemistry
Fixed effects	year*subject	indiv	year*subject	indiv	year*subject	indiv

	(7)	(8)	(9)	(10)	(11)	(12)
Girl	0.313**		0.144		0.284**	
	(0.152)		(0.149)		(0.142)	
Girl*Is	-0.952**	-0.932**	-0.574	-0.548	-0.828***	-0.855***
	(0.376)	(0.374)	(0.447)	(0.456)	(0.303)	(0.300)
Observations	2,081	2,081	1,668	1,668	4,938	4,938
R-squared	0.003	0.246	0.001	0.221	0.003	0.211
Track	Biology-Geology	Biology-Geology	Social Sciences	Social Sciences	Humanities	Humanities
Fixed effects	year*subject	indiv	year*subject	indiv	year*subject	indiv

*Note: The dependent variable is the candidate's difference between the oral and written test scores. "Is" is the subject feminization index. Robust Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table A6 :Distribution of the Estimated Gender Bias with indexes for Subjects and Tracks Degree of Feminization

Sample :Position wrt threshold	Below	Around	Above	Below	Around	Above
	(1)	(2)	(3)	(4)	(5)	(6)
Girl	0.383**	-0.123	0.206			
	(0.184)	(0.230)	(0.248)			
Girl*I_s				-0.759***	-0.746**	0.005
				(0.274)	(0.376)	(0.320)
Girl*I_j	-1.238**	0.218	-0.761			
	(0.497)	(0.611)	(0.657)			
Girl*I_{js}	-1.043***	-0.822***	-0.039			
	(0.275)	(0.302)	(0.288)			
Observations	5,246	3,812	3,764	5,246	3,812	3,764
R-squared	0.024	0.027	0.044	0.290	0.341	0.312
Track	all	all	all	all	all	all
Individual fixed effects	No	No	No	Yes	Yes	Yes
year*subject controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: The dependent variable is the candidate's difference between the oral and written test scores. Columns (2) and (5) give the results estimated on the 30% candidates who were "around" the admission threshold at the end of the eligibility step (15% above, 15% below). Estimates for candidates below and above the latters are presented respectively on columns (1)-(4) and columns (3)-(6). "Ij" is the subject feminization index, "It" the track feminization index and "Ij_t" their difference. Robust Standard errors in parentheses.
**** p<0.01, ** p<0.05, * p<0.1*