# Do gender quotas pass the test? Evidence from academic evaluations in Italy*

Manuel Bagues † Mauro Sylos-Labini ‡ Natalia Zinovyeva §

This version: August 29, 2014
First version: June 21, 2014

**Abstract**

This papers studies how the presence of women in academic committees affects the chances of success of male and female candidates. We use evidence from Italy, where candidates to Full and Associate Professor positions are required to qualify in a nation-wide evaluation known as *Abilitazione Scientifica Nazionale*. This evaluation was conducted between 2012 and 2014 in 184 academic disciplines and it attracted around 70,000 applications. In each field, committee members were selected from the pool of professors that had volunteered for the task using a random lottery. We estimate the causal effect of committees' gender composition on candidates' chances of success exploiting the existence of this system of random assignment. In a five-member committee, each additional female evaluator decreases by 2 percentage points the success rate of female candidates relative to male candidates. Information from 274,000 individual evaluation reports shows that, in mixed-gender committees, male and female evaluators are equally biased against female candidates, suggesting that the presence of women in the committee affects the voting behavior of male evaluators.

**Keywords:** gender quotas, discrimination, academic promotions.
**JEL Classification:** J71, J16, J45.

*We are grateful for all the useful comments and suggestions provided by David Cuberes, Cecilia Garcia-Peñalosa, Moreno Marzolla and participants at presentations at the *Festival dell'Economia* in Trento, Swedish Institute for Social Research (SOFI), the Workshop 'Gender in the Boardroom' (Sheffield), the 7th COSME-FEDEA Gender Workshop and the Annual Meeting of the French Economic Association in Lyon.

†Aalto University and IZA, Helsinki, Finland; email: manuel.bagues@aalto.fi
‡University of Pisa, Italy; email: mauro.syloslabini@sp.unipi.it
§Aalto University, Helsinki, Finland; email: natalia.zinovyeva@aalto.fi

# 1 Introduction

To achieve gender balance, academia has still a long way to go. In Europe women account for 46% of PhD graduates, 37% of associate professors and only a mere 20% of full professors (European Commission 2013). The US displays similar patterns and in Japan the gender imbalance is even larger (National Research Council 2009, Abe 2012).

There are several possible explanations for the persistent underrepresentation of women in top positions in academia. According to the pipeline theory, it would be mainly a matter of time that women would move their way through a metaphorical pipeline to reach top-level jobs. However, in most disciplines, the share of women among faculty members remains low even after decades of improved recruitment of women at the undergraduate and the doctoral level (Ginther and Kahn 2004, 2009). Gender differences in promotion rates might reflect differences in productivity, perhaps due to the existence of gendered roles at the household level or the lack of female mentors and role models (Blau et al. 2010). Furthermore, some authors have pointed out that women are less likely to apply for promotions (Bosquet, Combes and Garcia-Peñalosa 2013, De Paola, Ponzo and Scoppa 2014), maybe due to existence of gender differences in the preference for competitive environments (Niederle and Vesterlund 2007; Buser, Niederle and Oosterbeek 2014) or in bargaining abilities in the labour market (Babcock, Gelfand, Small and Stayn 2007; Blackaby, Booth and Frank 2005). Moreover, women seem to devote more time to tasks that are socially desirable but might not taken into account in promotion decisions (Vesterlund, Babcock and Weingart 2014).

Beyond these supply-side explanations, the paucity of women in top positions in academia has been also sometimes attributed to the existence of gender discrimination by the (mostly male) evaluators that decide on hiring and promotion decisions.[1] It has been argued that male evaluators might be subject to gender stereotypes.[2] They are also more likely to be acquainted with male candidates and these connections might perhaps affect the outcome of the evaluations (Zinovyeva and Bagues 2014). The lack of female evaluators might also be detrimental for female candidates if men and

[1] Gender discrimination in academia remains a controversial issue. According to a meta-analysis by Ceci and Williams (2011), the more recent empirical evidence fails to support assertions of discrimination in manuscript reviewing, interviewing, and hiring. However, other studies show that female researchers might still receive lower evaluations than male researchers with identical characteristics (e.g. Steinpreis, Anders and Ritzke 1999, Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman 2012).

[2] For instance, a report by a Spanish governmental organization, the Foundation for Science and Technology, claims that in "academia, promotion is based on a system [...] that benefits men more than women, since the barriers arise when mostly male committees evaluate female candidates and reject their promotion" (Fundación Española para la Ciencia y la Tecnología, *Mujer y Ciencia: La situación de las Mujeres Investigadoras en el Sistema Español de Ciencia y Tecnología*, (FECYT, 2005), page 48.

women differ in terms of which research areas are more valuable. For instance, several authors have documented the existence of gender segregation across different subfields in Economics (Dolado et al. 2012, Hale and Regev 2014). If evaluators overrate the importance of their own types of research, a preference for same-sex candidates might arise.

These arguments have reached policy makers. Gender quotas in scientific committees have been already introduced in countries such as Spain, Finland, Sweden, and Norway, and they are being considered in other countries. A recent report of the European Commission argues that a gender-mixed composition of scientic committees is needed to ensure that constant progress is made towards gender-equality in research and scientic careers.[3] However, the empirical evidence on the impact of committees' gender composition is scarce, typically based on small samples, and it is rather inconclusive. Sometimes, applicants seem to benefit from the presence of same-sex evaluators (Li 2012, De Paola and Scoppa 2014), but most often gender does not seem to play any role (Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman 2012; Steinpreis, Anders and Ritzke 1999; Abrevaya and Hamermesh 2012; Jayasinghe, Marsh and Bond 2003) or applicants obtain relatively better evaluations from opposite-sex evaluators (Broder 1993; Ellemers, Heuvel, de Gilder, Maass and Bonvini 2004). From a policy perspective, the lack of more extensive and conclusive evidence is disappointing. Gender quotas are costly for senior women, as they increase disproportionally the time that they have to devote to evaluation committees. A better understanding of its impact on recruitment and promotion decisions is key in order to determine whether they are cost-effective.

Zinovyeva and Bagues (2011) provide the first large-scale experimental evidence on the role of evaluators' gender in academic evaluations. They analyze the evaluations received by 30,000 applicants to Associate and Full Professor positions in all academic disciplines in Spain between 2002 and 2006, when the system known as *Sistema de Habilitación* was in place. The institutional set up provides an interesting randomised natural experiment. Within this system, applicants have to qualify in a centralized evaluation which is performed periodically by evaluation committees at the national level. Each time, committee members are selected by public officials from the corresponding pool of eligible evaluators using a random lottery. As a result, candidates with similar characteristics may, by the luck of the draw, be assigned to evaluation committees with different gender compositions. On average, the gender of evaluators does not have a statistically significant impact on the chances of success of male and female candidates. The authors also show that the impact of evaluators' gender does not depend on the degree of feminization of the field, it is similar across different

[3]European Commission, *She Figures 2012: Gender in Research and Innovation* (Luxembourg: Publication Office of the European Union, 2013), page 7.

disciplinary groups (Life Sciences, Natural Sciences, Engineering, Social Sciences and Humanities), but it seems to vary across different types of positions. Evaluators exhibit a same sex preference in competitions to Full Professor, but an opposite sex preference in competitions to Associate Professor.

In this paper, we analyze how the gender composition of academic committees affects evaluations using evidence from another large-scale natural randomized experiment: qualification evaluations in Italy. This system of promotions, known in Italian as *Abilitazione Scientifica Nazionale*, is largely similar to the system that was in place in Spain between 2002 and 2006, but also presents some distinctive features. As in the Spanish case, candidates to Associate and Full Professor positions are required to qualify first in an evaluation performed at the national level by a committee whose members are selected randomly from a pool of eligible evaluators in the field. However, while in Spain evaluations involved a research seminar given by the candidate, in Italy the evaluation relied completely on CVs and publications and it does not require any personal interaction between evaluators and candidates. Most importantly, the procedure was extremely transparent. All the relevant information — including candidates' and evaluators' CVs, bibliometric indicators and individual evaluation reports — was publicized online.[4]

Our database includes information on 69,020 applications in 184 different fields. Compared to male applicants, female applicants tend to exhibit a worse CV in terms of the quantity and quality of publications and a lower unconditional success rate. However, we do not observe any significant differences in their conditional success rate: on average, women are as likely to qualify as male candidates with similar observable research production. Next, we examine whether the chances of success of male and female candidates depend on the gender composition of committees. We compare the success rate of male and female candidates in exams where the gender composition of the committee is expected to be similar but, due to the random draw, committees end up having a different gender mix. As expected, these groups of candidates are statistically similar in every observable predetermined dimension. At the same time, we find that the success rate of female applicants is significantly lower whenever, due to the random draw, the committee includes more female evaluators. In a five-member committee, each additional female evaluator decreases by around two percentage points (five per cent) the success rate of female applicants relative to the success rate of male applicants. This effect is mainly driven by evaluations in exams for Associate Professor positions.

The availability of very detailed data also allows to examine more closely the po-

[4] A national evaluation agency collected and publicized information on bibliometric indicators for all candidates. Committees are not obliged to use this information, but there is a clear 'nudge' from the authorities recommending that committees take them it into account.

tential mechanisms at work. In our data approximately 14% of candidates voluntarily withdraw from the evaluation process when the lottery is drawn and the identity of evaluators is known. We find that this decision does not depend (significantly) on the gender composition of committees. Hence, the gender composition of committees matters because it affects evaluations, and not because it influences candidates' final decision to apply.

We also analyze the individual voting behavior of committee members. Committee members typically collect information about the quality of candidates, they share it, debate about it and take a vote. Thus, the presence of women in the committee might affect the final outcome directly through their own vote, but also indirectly through their impact on the voting decisions of other committee members. Information from around 274,000 individual voting reports shows that, in mixed-gender committees, male and female evaluators are equally biased against women, suggesting that the presence of women in the committee affects the voting behavior of male evaluators. We also examine the length of evaluation reports. This variable might partly reflect the amount of information that was available about the quality of a candidate. Male and female evaluators write reports of similar length independently of the gender of the applicant.

Overall, the evidence in this paper suggests that, on average, academic evaluators do not exhibit a preference for same-sex candidates. If anything, female candidates tend to be more successful in committees with fewer female evaluators. The potential introduction of gender quotas in these committees would be detrimental both for junior female researchers, who would have lower chances of being promoted, and for senior ones, who would have less time available to work on their research.

The structure of the paper is as follows. First, section 2 discusses the related empirical literature. Section 3 describes the institutional details of the system of *Abilitazione Scientifica Nazionale*. Next, in section 4 we present the data and in section 5 we analyze the role of candidates' and evaluators' gender in evaluations and we explore possible explanations for the observed evidence. Finally, in section 6 we summarise our results and we examine some possible policy implications.

# 2 Literature review

The role of the evaluators' gender in academic evaluations has been studied by a number of authors.[5] Typically the main empirical challenge has been to identify variations in the gender composition of committees that are not somehow related to the relative quality of male and female candidates. At least two papers provide experimental evidence. Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman (2012) performed

[5] A related literature also analyses the role of evaluators' gender in non academic occupations (e.g. Bagues and Esteve-Volart 2010) or in sport activities (e.g. Sandberg 2014).

a randomized double-blind study where 127 biology, chemistry, and physics professors from research-intensive universities rated the application materials of an undergraduate science student who had (ostensibly) applied for a science laboratory manager position. Both male and female faculty judged a female student to be less competent and less worthy of being hired than an identical male student. Steinpreis, Anders and Ritzke (1999) asked 238 male and female academic psychologists to review the CV of an applicant who was assigned randomly a male or a female name. They do not observe any impact of evaluators' gender on the assessments.

Another empirical strategy is to compare the assessments received by the same application from different independent evaluators. The key assumption in this case is that the gender of candidates does not affect the assignment of evaluators.[6] Broder (1993) analyzes 1,479 grant proposals to the Economics program of the National Science Foundation. She finds that female reviewers rate female-authored papers lower than do their male colleagues. Abrevaya and Hamermesh (2012) study the referee evaluations received by 2,940 submissions to a leading field journal in Economics. They find no interaction between the referees' and authors' gender. Ellemers, Heuvel, de Gilder, Maass and Bonvini (2004) study the assessments by faculty members of the work commitment of 212 PhD students in the Netherlands and in Italy. Female faculty members are most inclined to hold stereotypical views. Jayasinghe, Marsh and Bond (2003) analyze the ratings obtained by 687 grant proposals at the Australian Research Council. They do not find any evidence suggesting that assessors favor same-sex or opposite sex researchers.

While the above literature considers the behavior of independent individual evaluators, very often hiring and promotion decisions are taken by committees composed of several members. A few articles have studied the role of committees' gender composition. For instance, Li (2012) examines the extent to which evaluators' gender influences peer review at the US National Institutes of Health, using information from around 20,000 successful applications. Taking into account the observable information about the past and future research performance of successful candidates, Li concludes that reviewers exhibit a same-sex preference when they review competing applications, but not when they evaluate new applications. De Paola and Scoppa (2014) study how variations in the gender composition of evaluation committees affect the chances of success of 1,000 applicants to Full and Associate Professor positions in Economics and Chemistry in Italy, documenting a same sex preference.

Zinovyeva and Bagues (2011) provide the first evidence from a large-scale randomized natural experiment involving all academic disciplines in a given country. They analyze the evaluations received by 30,000 applicants for Full and Associate Professor

[6]For instance, it would be problematic if specially demanding evaluators are assigned to evaluate same-sex (or opposite-sex) candidates.

positions in every academic field in Spain between 2002 and 2006. On average, the chances of success of female and male candidates are unaffected by the gender composition of committees, but there might be some heterogeneous effects depending on the position at stake. Similarly, in this paper we contribute to the literature providing evidence from another large-scale randomized natural experiment. In this case, we use information from qualifying exams for Associate and Full Professor positions that were held between 2012 and 2014 in every discipline in Italy. These evaluations involved 70,000 candidates. As in Zinovyeva and Bagues (2011), the institutional setup provides a very transparent source of exogenous variation in committee composition: the random lottery that selects committee members out of the pool of eligible evaluators. Moreover, as we explain in detail below, the availability of very detailed information about committee members' individual voting behavior allows us to shed some light on the interactions that arise within committees.

# 3 Background

Most Italian universities are public and the recruitment of full and associate professors is regulated by national laws.[7] Before 2010, recruitment procedures were managed locally by each university. According to some authors, this system fostered nepotism (Durante et al. 2011). In 2010, a two-stage procedure similar to those already in place in other European countries was introduced (e.g. France and Spain).[8]

In the first stage, candidates to Associate Professor (AP) and Full Professor (FP) positions are required to qualify in a national-level exam. Evaluations are conducted separately in 184 scientific fields (in Italian named *settore concorsuale*) designed by the Ministry of Education. A positive evaluation is valid for four years while a negative one implies the exclusion to participate in further national evaluations during the following two years. Qualified candidates can participate in the second stage, which is managed locally by each university. This may consist in either an open competition or the assessment of an internal candidate.[9]

[7] According to *OECD Education at a glance* (2013 edition), in 2011 about 92% of students in tertiary education were enrolled in 66 public universities and the remaining 8% in 29 independent private institutions.

[8] Law number 240/2010, also known as "Gelmini reform" after the name of the minister of Education.

[9] From 2017 on this second option will be viable only for assistant professors who want to become associate.

## 3.1 The national examination

The first national evaluation, which we study in this paper, was performed between 2012 and 2014, about two years after the reform approval.[10] The time structure of the examinations was as follows. The call for eligible evaluators was published in June 2012. The deadline to apply was August 28. Once the list of eligible evaluators was settled, the Ministry publicized their identities and their CVs. In the meantime, the call for candidates' applications was issued in July. Candidates could apply to multiple fields and positions. The deadline for applications was November 20. Then, in January 2013, when the initial list of applicants was settled but before it was publicized on-line, committee members were selected by random draw.

Soon afterward, and before having access to the list of candidates, each evaluation committee is required to draft and publish on-line a document describing the general criteria that would be used to grant a positive evaluation. For candidates, the deadline to withdraw applications expires two weeks after these criteria are publicized. Finally, evaluation committees are informed about the final list of candidates and the examination takes place. Below we explain in more detail how evaluators are selected and how the evaluation is performed.

### 3.1.1 Selection of committees

Evaluation committees include five members selected by a random lottery. Four members are drawn from a pool of eligible evaluators affiliated to an Italian university (hereafter 'Italian') and the remaining member is selected from a pool of eligible evaluators who are affiliated to a foreign university (hereafter 'foreign').[11] The only constraint to the randomization process is that no university can have more than one evaluator within a single committee. All evaluators are in charge for two years. In case an evaluator resigns, a replacement is selected randomly from the corresponding list of eligible evaluators.

The set of 'Italian' eligible evaluators includes all full professors in the field who volunteered for the task and satisfied some minimum quality requirements. In Math, the Natural and Life Sciences, and Engineering eligible evaluators are required to have a research production above the median of full professors in the field in at least two of the following three dimensions: (i) the number of articles published in scientific journals, (ii) the number of citations, (iii) and the H-index.[12] In the Social Sciences

[10] A detailed description of the process is available at `http://abilitazione.miur.it/public/index.php?lang=eng`

[11] Whenever the pool of foreign professors included less than four professors, all five committee members were drawn from the pool of eligible evaluators based in Italy.

[12] More precisely, this includes Mathematics and IT, Physics, Chemistry, Earth Sciences, Biology, Medicine, Agricultural and Veterinary Sciences, Civil Engineering and Architecture (with the exception of Design, Architectural and Urban design, Drawing, Architectural Restoration, and Urban and

and the Humanities, the research performance of eligible evaluators has to be above the median in at least one of the following three dimensions: (i) the number of articles published in high quality scientific journals (in what follows, 'A' journals),[13] (ii) the overall number of articles published in any scientific journals and book chapters, and (iii) the number of published books.

The set of 'foreign' eligible evaluators includes Full Professors affiliated to a university from an OECD country who volunteered for the task. 'Foreign' eligible evaluators have to satisfy the same research requirements as 'Italian' ones. While 'Italian' evaluators work *pro bono*, OECD evaluators receive €16,000 for their participation in two evaluations.

### 3.1.2 The evaluation

A positive assessment of each candidate requires a qualified majority of four votes. Committees make their assessments based on candidates' CVs. The evaluation does not involve neither exams nor interviews. In principle, committees have full autonomy on the criteria to be used in the evaluation. Nonetheless, an independent evaluation agency appointed by the ministry collected and publicized information on the research productivity of candidates in the previous ten years, as measured by the three bibliometric indicators and asked the committees to evaluate candidates' publications also with reference to the above bibliometric indicators. These productivity measures are normalized taking into account time since first publication and job interruptions, typically related to maternity or paternity.

At the end of the process committees release for each candidate a (i) collective evaluation explaining how their final decision was taken and (ii) five individual evaluations explaining each evaluators' position (see Figure 1).

# 4 Data

We use information from the first edition of the *Abilitazione Scientifica Nazionale*. For each of the 184 academic fields, the dataset includes information on (i) eligible evaluators; (ii) evaluators that eventually served in committees, (iii) applicants, and (iv) the final outcome of the evaluation. Below we describe in detail the database.[14]

Regional Planning), Industrial and Information Engineering, and Psychology.

[13] An evaluation agency determined with the help of several scientific committees the set of journals to be considered as high quality in each field.

[14] We collected the CVs of candidates and evaluators and their final evaluations from the webpage of the Ministry of Education. To avoid homonymity problems, we have excluded 14 candidates that had the same name and surname within the same exam.

## 4.1 Evaluators

Overall 7,239 FPs volunteered to serve in committees. In the average field, the pool of eligible evaluators includes 32 'Italian' FPs and eight 'foreign' FPs. Approximately 20% of 'Italian' evaluators are women (Table 1, upper panel). The 'foreign' pool is less feminized: only 12% of 'foreign' evaluators are women. Four out of the five committee members were selected through a random draw out of the pool of 'Italian' eligible evaluators and the remaining committee member was drawn out of the pool of 'foreign' eligible evaluators.[15] Taking into account the composition of both pools, the expected share of women in the committee is around 18%.[16] The actual gender composition of committees is similar to the expected one: on average 18% of committee members were women. Approximately one out of every thirteen evaluators resigned and was replaced by another eligible evaluator. These replacements increased slightly the share of women in committees, but the difference is not statistically significant (Table 1, lower panel).

## 4.2 Initial set of applicants

The initial number of applications was equal to 69,020. On average, there were 375 applications per field, 117 in examinations for Full Professor (FP) positions and 258 in examinations for Associate Professor (AP) positions. Several candidates participated in more than one evaluation, either in different fields or in different categories of the same field. The average candidate participated in 1.5 examinations.

As shown in Table 2, most applicants are are affiliated to an Italian institution (96%) and, not surprisingly, applicants tend to be older in competitions for a position of FP than in competitions for a position of AP (49 vs. 43 years old). In qualification exams for positions of FP around 31% of candidates are women and, among candidates for positions of AP, the degree of feminization is slightly larger: women account for 41% of candidates. Applicants to FP are more likely to hold a permanent position in an Italian university (74% vs. 47%) and, out of these candidates, three fourths hold a position in the same field to which they are applying. We also collected information on the relative order of application for all candidates. We normalized this variable uniformly between 0 and 1. In principle, the timing of the application can reflect both candidates' self-confidence and their quality.

CVs include detailed information on candidates' research production. The average

[15] In twenty fields the pool of foreign evaluators did not reach the required threshold of four members. In these cases all five committee members were drawn from the 'Italian' pool. Additionally, in two fields the pool of foreign evaluators included originally four members, but after the resignation of the foreign member her replacement was selected from the 'Italian' pool.

[16] We have calculated the expected gender composition of committees using a simulation with 1000 draws, taking into account that the lottery that decided committee composition was subject to the constraint that committees cannot include more than one member from the same university.

CV has 16 pages and it reports around 64 research outputs, including journal articles (36), books (2), book chapters (7), and patents (0.24). A typical paper is coauthored by 5 authors and only 34% of papers are single authored. The candidate reports to be the first author in 22% of the occasions and the last author in 12%.[17] We also gathered information on two variables that proxy the quality of publications. In the Social Sciences and in the Humanities, we collected information on the total number of articles published in A journals.[18] Candidates in these fields published on average 4 A-journal articles. In Sciences, we also consider the *Article Influence Score* (AIS) of journals.[19] Summing up the AIS of all the publications of a candidate, the total AIS of the average candidate is around 61. Not surprisingly, candidates to FP positions have a relatively stronger publication record. They have a longer publication list (89 vs 53 publications) and their total AIS is also significantly higher (86 vs 49).

Table 2 also provides information about candidates' characteristics, by gender (columns 5-7). Female applicants tend to display slightly longer CVs but their publication record is significantly weaker (Table 2, columns 5-7). They are also younger, less likely to be based abroad, and more likely to have a permanent contract in the same field where they applied. We do not observe any significant difference in the timing of the application between men and women.

## 4.3 Final set of applicants

Approximately 14% of applicants (9,870 out of 69,020 applications) withdrew their application once the identity and the criteria of evaluators were made public. The incidence of withdrawals was significantly larger in FP examinations than in AP examinations (16% vs 13%). Withdrawals were also more common among female applicants (17% vs 13%).

The evaluation agency of the Ministry of Education collected and publicized detailed information about the research production during the 10 previous years of the final set of candidates. In the Social Sciences and the Humanities, it includes the number of books, the total number of academic articles, and the number of articles published in A-journals. In Scientific fields, the evaluation agency collected information on candidates' number of publications, citations received and on their H-index.

Using these bibliometric indicators, the evaluation agency compared the research productivity of candidates with the research productivity of professors in the category to which they applied. Around 38% of candidates were above the median in each of

[17] Some candidates list their name first among the list of authors of a given paper regardless of the actual ordering of authors in the publication.

[18] We define A-journals following the criteria of the Italian Evaluation Agency (ANVUR).

[19] This indicator is available for all publications in the Thomson Reuters Web of Knowledge. It is related to *Impact Factor*, but it takes into account the quality of the citing journals, the propensity to cite across journals and it excludes self-citations.

the three corresponding bibliometric dimensions. Applicants in FP examinations were relatively stronger than in AP examinations (42% vs. 36% are above the median) as well as male applicants (40% vs 34%).

Around 43% of the applicants that did not withdraw their application managed to qualify. The success rate is slightly lower if we consider all candidates that initially applied (37%). In addition to the final decision of the committee, we also collected information on the length of evaluation reports and on the voting behavior of each committee member.[20] The average collective evaluation report included 280 words, while individual reports were slightly shorter, around 176 words. 43% of these reports were favorable to the candidate and most of the time (80%) decisions were taken by unanimity.

Success is strongly correlated with observable research productivity (Figure 2). Among candidates whose quality was below the median in every dimension, only 4% managed to succeed. On the contrary, 63% of candidates that excelled in all three dimensions qualified.

# 5 Empirical analysis

The empirical analysis is organized as follows. First, we provide some descriptive information about the presence of women along the academic career ladder. Second, we compare the evaluations received by male and female applicants that participated in the *Abilitazione Scientifica Nazionale* and examine whether the gender gap varies with the gender composition of committees. Finally, in order to get a better understanding of the observed patterns, we study the impact of committees' gender composition on candidates' decision to withdraw their applications and we also analyze the information provided by individual voting reports.

## 5.1 Gender gap in promotions

The proportion of women decreases along the academic career ladder. As shown in Table 3, approximately 45% of assistant professors affiliated to Italian universities are women.[21] The share of women among Assistant Professors applying for positions of Associate Professor in the national qualification exams is slightly lower, around 43%,

[20] We conducted a text analysis of the individual evaluation reports. We identified approximately 9,000 different sentences that indicate the evaluator's decision to fail or to pass a given candidate. This sentences were used in approximately 274,000 of the 295,000 available individual evaluation reports.

[21] We collected from the Ministry of Education information on all assistant and associate professors in the field who were affiliated to an Italian university in December 2012, shortly before *Abilitazione Scientifica Nazionale* took place. Unfortunately, we do not have systematic information on the identity of other potential applicants, such as Italian researchers based abroad, researchers based in an Italian research centre (CNR) or potential applicants based in a Italian university who had some alternative contract (i.e. postdoctoral scholarship).

and women constitute only 40% of assistant professors that qualified. Similarly, while 35% of associate professors based in Italian universities are women, in evaluations for Full Professor positions the share of women is equal to 32%, and it is slightly lower among those who qualify, around 30%.

Next, we analyze why female candidates are underrepresented among successful candidates, paying particular attention to the role of committees' gender composition.

## 5.2 Gender differences in performance

Overall, male candidates tend to be more successful. The (unconditional) gender gap is equal to 2.2 percentage points (Table 4, column 1). In order to disentangle how much this reflects differences in the quality of candidates or in the assessments of evaluators, we compare the chances of success of male and female candidates taking into account their observable research productivity and the field and category where the evaluation is being performed. In particular, we estimate the following equation:

$$Y_{i,e} = \beta_0 + \beta_1 Female_i + \mathbf{X_i}\beta_2 + \mathbf{Z_i}\beta_3 + \mu_e + \epsilon_{i,e} \quad (1)$$

where $Y_{i,e}$ is a dummy variable that takes value one if candidate $i$ qualifies in evaluation $e$ (for instance, evaluation for the position of Associate Professor in Applied Economics) and value zero if the candidate fails to qualify. $Female_i$ is a dummy variable indicating the gender of candidate $i$. We control for the observable productivity of candidates using the information described in section 4.2 and, to account for possible non linearities, we also consider a set of indicator dummies that take value one if candidate's quality in a certain dimension is above the median in the corresponding category and field ($\mathbf{X_i}$). In addition to direct measures of research productivity, we also take into account a number of individual characteristics that might be correlated with quality, such as candidate's age, country of residence or the relative order of application ($\mathbf{Z_i}$). In the case of candidates based in an Italian university, we also consider their affiliation, the type of contract and whether they hold this contract in the same academic field to which they are applying. A set of dummies at the exam level ($\mu_e$) controls for any overall differences across exams that might affect the success rate of male and female candidates in a similar way. As a result, the coefficient $\beta_1$ captures any systematic differences in the success rate of female and male applicants that cannot be explained by candidates' observable characteristics. We cluster standard errors at the field level to account for the fact that the same committee evaluates all candidates within a field.

Most of the unconditional gender gap is explained by differences in candidates' observable research productivity. The residual gender gap is about 0.4 percentage points and it is not significantly different from zero (columns 2 and 3). We also analyze how the gender gap varies depending on the category or the type of field. The gender

gap is slightly larger in competitions to Associate Professor and in the Social Sciences and the Humanities, but it is not significantly different from zero at standard levels (columns 4 and 7).

## 5.3 Does the gender composition of committees matter?

On average, female candidates are as likely to qualify as male candidates with similar observable research productivity. Next, we investigate whether the presence of female evaluators in the committee makes any difference. We estimate again equation (1) including now the interaction between the gender of the candidate and the share of women among evaluators. Interestingly, while in committees where all evaluators are male there is no gender gap, in mixed gender committees there is a significant gap in favor of male candidates (column 1, Table 5). This effect is slightly larger and more precisely estimated in evaluations of application for AP positions (columns 2 and 3). The magnitude is relatively similar in different disciplinary groups (columns 4 and 5).

The above estimates cannot be easily interpreted as causal effects. They have to be considered with caution due to the potential existence of differences in the (unobservable) quality of male and female candidates across different fields. To address this endogeneity problem, we examine how the gender composition of committees affects male and female candidates' chances of success, exploiting the exogenous variation in committee composition caused by the random draw. More precisely, we compare the success rate of male and female candidates who initially were expected to face an evaluation committee with the same gender composition but, due to the random draw, are assigned to committees with different gender compositions. To avoid any potential selection bias, we consider the pool of initial applicants, independently of whether they withdrew their application when they received information about committee composition. We estimate the following equation:

$$\begin{aligned} Y_{i,e} = \beta_0 + \beta_1 Female_i + \beta_2 Female_i * Female_e + \\ + \beta_3 Female_i * Female_e^{expected} + \mathbf{X_i}\beta_4 + \mathbf{Z_i}\beta_5 + \mu_e + \epsilon_{i,e} \end{aligned} \tag{2}$$

where $Y_{i,e}$ takes value one if the candidate qualifies and value zero in the candidate withdrew her application or failed the evaluation. The variable $Female_e$ represents the share of female evaluators in the committee that was initially randomly drawn, before any evaluator resigned, and $Female_e^{expected}$ is the expected share of women in the committee. In order to increase the accuracy of the estimation, we also include information about individual observable productivity and individual characteristics that might be correlated with quality. Standard errors are clustered at the field level.

The coefficient $\beta_2$ captures the causal effect of committees' initial gender composition on the success rate of female candidates, relative to male candidates. The key

identifying assumption is that, conditional on the expected composition of the committee, its actual composition is uncorrelated with any unobserved factor that might affect candidates' evaluation. In other words, the consistency of $\beta_2$ relies on the assignment being indeed random. The way in which the randomization was conducted – selecting a random sequence of numbers that was then applied to several fields – suggests that there was little room for manipulation. Nonetheless, we explicitly test the randomness of the assignment. As expected, there exists no significant correlation between the random shock to committee composition and any predetermined variable (Table 6).

Table 7 reports the estimates from equation 2. The initial share of women in the committee has a significant negative impact on the relative chances of success of female candidates (column 1). In column 2, we control for candidates' observable research productivity and individual characteristics. The point estimate is slightly larger but, as expected, statistically similar. In column 3, we take into account the fact that some evaluators declined to participate in committees. More precisely, we instrument the final gender composition of the committee using the initial composition of the committee determined by the random draw. Results are very similar, if anything the impact of committees' gender composition is slightly larger. In terms of economic magnitude, an additional woman in a five-member committee decreases the relative chances of success of female candidates by approximately two percentage points ($\Delta$ $Female_e$=1/5; $\beta_2 = 0.123$). This pattern seems to be driven mainly by the evaluations of applications for AP positions and the magnitude of the effect is very similar in Sciences and in Social Sciences and Humanities (columns 4-7).

## 5.4 Do female applicants shy away from male evaluators?

The presence of women in the committee has a negative impact on the chances of success of female candidates. This might reflect evaluators' assessments or, alternatively, it might be that the gender composition of committees affects differently male and female candidates' decision to participate in the evaluation process. Some candidates might withdraw their application if they believe that evaluators might have some (direct or indirect) preference for same-sex or opposite-sex candidates. The composition of committees can also affect the information that is available to male and female candidates about their own chances of success. Evaluators might potentially provide more accurate information to same-sex candidates.

Approximately 13% of male candidates withdraw their application before the evaluation is conducted. The proportion of women that decides not to proceed with their application is significantly higher, approximately 17%. However, these gender differences in attrition rates are not related to the gender composition of committees (Table 8, columns 1-3). Women are slightly more likely to withdraw their application when

there are women in the committee, but the coefficient is economically small and it is not statistically significant. A similar result arises when we consider separately exams to FP and exams to AP positions, or exams in the Social Sciences and the Humanities vs. exams in the Sciences (columns 3-8).

## 5.5 Individual evaluations

Committees that include female evaluators tend to be less favorable towards female candidates. In principle, it is unclear whether this effect is directly driven by the voting behavior of female evaluators or, perhaps, by the reaction of male evaluators to the presence of female evaluators in the committee. In order to shed light on this issue we analyze the individual voting behavior of each committee member. We run the following equation at the level of individual evaluation:

$$\begin{aligned} Y_{i,c} = \beta_0 + \beta_1 Female_i + \beta_2 Female_i * Female_e + \\ + \beta_3 Female_i * Female_e^{expected} + \mathbf{X_i}\beta_4 + \mathbf{Z_i}\beta_5 + \mu_e + \epsilon_{i,c} \end{aligned} \tag{3}$$

where the dependent variable $Y_{i,c}$ takes value one if committee member $c$ voted in favor of candidate $i$ and the set of independent variables is similar to the one consider in equation 2. In all-male committees, female candidates tend to obtain a similar number of positive votes as equally qualified male candidates, but in gender-mixed ones women have a significantly worse performance, although in this case the gap is only significant at the 10% (Table 9, columns 1 and 2).

Two possible mechanisms might explain this pattern. First, female evaluators might be biased against female candidates. That would explain why the presence of women in the committee hurts the chances of success of female applicants. Second, maybe male evaluators change their voting behavior in the presence of women in the committee, becoming tougher with female candidates. To disentangle between the two mechanisms, we examine whether male and female evaluators vote differently depending on the gender of the candidate. In order to account for potential differences in the quality of candidates that might be observed by evaluators but not by the econometrician, we run the following equation:

$$Y_{i,c} = \beta_0 + \beta_1 Female_i * Female_c + \mu_i + \epsilon_{ic} \tag{4}$$

where candidate fixed effects control for any characteristics of candidates that are observable by all committee members and the error term $\epsilon_{ic}$ captures any remaining differences in the evaluation criteria of committee members, for a given candidate.

If anything, female evaluators are slightly more favorable towards female candidates than male evaluators, but the difference is not statistically significant (colunm 3). The

gender gap observed in mixed-gender committees cannot be explained exclusively by the behavior of female evaluators. Both of male and female evaluators sitting in mixed-gender committees are relatively less favorable towards female candidates.[22]

Finally, we examine the length of collective and individual evaluation reports. As we point out in a companion paper, evaluation reports tend to be longer when evaluators and candidates work in the same institution, they are co-authors or they do research in the same subfield (Bagues, Sylos-Labini and Zinovyeva 2014). The length of reports is also decreasing with the number of candidates that a committee has to evaluate, suggesting that committees with many candidates spend less time on each evaluation. In sum, the length of evaluation reports might reflect the amount of information that was available to evaluators about the quality of the candidate at the time. We run equation (3) using as dependent variable the length of committees' collective report. Evaluators write slightly longer reports on female candidates, but this gap is only significant at the 10% and it does not vary with the gender composition of the committee (Table 10, column 1). The length of individual reports does not depend either on the gender of the candidate (columns 2 and 3). In sum, there is no evidence suggesting that male or female evaluators have different information about the quality of male and female candidates.

# 6 Conclusions

In this paper, we analyze how the gender composition of evaluation committees affects the chances of success of candidates. We use the exceptional evidence provided by qualification evaluations to Full and Associate Professor positions in Italy. These competitions involved around 70,000 applicants and 1,000 evaluators and, due to their institutional design, they provide an extraordinary large-scale natural randomized experiment. The composition of evaluation committees is decided using a random lottery. As a result, groups of male and female candidates with similar characteristics face committees with different gender compositions.

On average, women are as likely to qualify than equally productive men. However, this gender gap varies significantly depending on the gender composition of evaluation committees. The presence of female evaluators in the committee has a significant negative impact on the success rate of female and male candidates. Each additional woman in the committee decreases the chances of success of female applicants by two percentage points relative to male applicants. Our analysis also shows that this effect is due to evaluators' behaviour, and not to candidates' potential decision to

[22]This information should be considered with certain caution, given that committee members share information and discuss each case before reaching a final decision. Another potential source of concern is that, sometimes, committee members who are in a minority might decide strategically to report a positive vote in their evaluation report, in order not to antagonize candidates.

withdraw from the competition. Information from individual votes within committees suggests that this bias partly reflects the differential behavior of male evaluators when women are present in the committee. Nonetheless, we are unable to provide a clear-cut explanation of why committees including female evaluators are relatively tougher with female candidates.

Our results have important policy implications for the introduction of gender quotas in evaluation committees in academia. These quotas have been introduced in several countries and they are intended to improve meritocracy and increase the presence of women in the upper echelons of the academic career. One possible justification would be the that all-male committees are biased in favor of male candidates (or the type of research done by men). Our results are not consistent with this hypothesis. In fact, mixed gender committees are less favorable towards women than all-male committees.

There are certain features of quotas that are not captured by our analysis. Evaluators that are explicitly chosen to represent a minority might behave differently, perhaps being more inclined to look positively at candidates belonging to their own group. Moreover, maybe once quotas are introduced, the strategic incentives of evaluators can be affected. For instance, given the disproportionate time that senior women would have to spend sitting in committees, there might be an incentive to increase the number of women that qualify. Keeping in mind these limitations, our results suggest that the introduction of gender quotas in evaluation committees might have unintended consequences. Quotas might be detrimental both for senior female researchers, who would have to spend more time sitting in committees, and for junior ones, who might have lower chances of success. A back of the envelope calculation suggests that a rule requiring the presence of at least 40% of women (or men) would imply that 700 less women qualify. Moreover, female full professors would have to form part of committees at least 3 times more often than men, and they would probably end up having less time to devote to their own research than their male colleagues.[23]

[23]We have chosen the 40% threshold following the example of countries such as Spain or Norway. In our data, only 18% of eligible evaluators are women. In order to satisfy a 40% quota, this implies that women would have to participate in committees at least 3 times as often as men [(0.4/0.18)/(0.6/0.82)]. On the other hand, 26,275 female candidates participated in these qualification exams. According to the estimate in Table 7, column 3, with a 40% quota 711 additional women would have failed to qualify [-0.123*(0.40-0.18)*26,275].

# References


Abrevaya, Jason and Daniel S. Hamermesh (2012), "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?," *Review of Economic and Statistics*, 94(1), 202-207.

Abe, Yukiko (2012), "The Academic Labor Market in Japan and the Presence of Women," *CSWEP Newsletter*, Fall, 9-10.

Babcock, Linda, Michele Gelfand, Deborah Small, and Heidi Stayn (2007), "Who gets to the bargaining table: The influence of gender and framing on the propensity to initiate negotiations," *Journal of Personality and Social Psychology*, 93, 600-613.

Bagues, Manuel and Berta Esteve-Volart (2010), "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *Review of Economic Studies*, 77(4), 1301-28.

Bagues, Manuel, Mauro Sylos-Labini and Natalia Zinovyeva (2014) "Can Transparency Overcome Evaluation Biases in Academia? Evidence from Italy", Aalto University, mimeo.

Blackaby, David, Alison L. Booth and Jeff Frank (2005), "Outside Offers And The Gender Pay Gap: Empirical Evidence From the UK Academic Labour Market," *Economic Journal*, 115(501), F81-F107

Blau, Francine D., Janet M. Curie, Rachel T.A. Croson and Donna K. Ginther (2010), "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial," *American Economic Review*, 100(2), 348-352.

Bosquet, Clément, Pierre-Philippe Combes and Cecilia Garcia-Peñalosa (2013), "Gender and Competition: Evidence from Academic Promotions in France", *Sciences Po Economics Discussion Papers* 17.

Broder, Ivy E. (1993), "Review of NSF Economics Proposals: Gender and Institutional Patterns," *American Economic Review*, 83(4), 964-970.

Buser, Thomas, Muriel Niederle and Hessel Oosterbeek (2014), "Gender, Competitiveness and Career Choices," *Quarterly Journal of Economics*, forthcoming.

Ceci, Stephen J. and Williams, Wendy M. (2011), "Understanding Current Causes of Women's Underrepresentation in Science," *Proceedings of the National Academy of Sciences*, 108.

CSWEP (2008), "Report of the Committee on the Status of Women in the Economics Profession 2008," *American Economic Review: Papers & Proceedings*, 99(2), 692-699.

De Paola, Maria, Michela Ponzo and Vincenzo Scoppa (2014), "Gender Differences in Attitudes Towards Competition: Evidence from the Italian Scientific Qualification," University of Calabria, mimeo.

De Paola, Maria and Vincenzo Scoppa (2014), "Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy," *Economica*, *forthcoming*.

Dolado, Juan Jose, Florentino Felgueroso and Miguel Almunia (2012), "Are Men and Women-Economists Evenly Distributed across Research Fields? Some New

Empirical Evidence," *Journal of the Spanish Economic Association*, 3(3), 367-393.

Durante, Ruben, Giovanna Labartino and Roberto Perotti (2011), "Academic Dynasties: Decentralization and Familism in the Italian Academia," *NBER working paper* n. 17572.

Ellemers, Naomi, van den Heuvel, Henriette, de Gilder, Dick, Maass, Anne and Alessandra Bonvini (2004), "The underrepresentation of women in science: Differential commitment or the queen bee syndrome?" *British Journal of Social Psychology*, 43, 1-24.

Ginther, Donna K. and Shulamit Kahn (2004), "Women in Economics: Moving Up or Falling Off the Academic Ladder," *Journal of Economic Perspectives*, 18(3), 193-214.

Ginther, Donna K. and Shulamit Kahn (2009), "Does Science Promote Women? Evidence from Academia 1973-2001," NBER chapters in: *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, 163-194.

Hale, Galina and Tali Regev (2014), "Gender Ratios at Top PhD Programs in Economics," *Economics of Education Review*, forthcoming.

Jayasinghe, Upali W., Marsh, Herbert W. and Nigel Bond (2003), "A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings", *Journal of the Royal Statistical Society*, 166(3), 279-300.

Li, Danielle (2012), "Gender bias in NIH Peer Review: Does it Exist and Does it Matter?", MIT, mimeo

National Research Council (2009), *Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty*, Washington D.C.: The National Academies Press.

Niederle, Muriel and Lise Vesterlund (2007), "Do Women Shy Away From Competition? Do Men Compete Too Much?", *Quarterly Journal of Economics*, 122(3), 1067-1101.

Moss-Racusin, Corinne A., John F. Dovidio, Victoria L Brescoll, Mark J. Graham and Jo Handelsman (2012), "Science faculty's subtle gender biases favor male students," *Proceedings of the National Academy of Sciences*, 109(41), 16474-16479.

Sandberg, Anna (2014), "Competing biases: Effects of gender and nationality in sports judging," Stockholm School of Economics, mimeo.

Steinpreis, Rhea E., Anders, Katie A. and Dawn Ritzke (1999) "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study," *Sex Roles*, 41(7), 509-527.

Vesterlund, Lise, Babcock, Linda and Laurie Weingart (2014), "Breaking the Glass Ceiling with "No": Gender Differences in Declining Requests for Non Promotable Tasks", University of Pittsburgh, mimeo.

Zinovyeva, Natalia and Manuel Bagues (2011), "Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment," IZA Discussion Paper 5537.

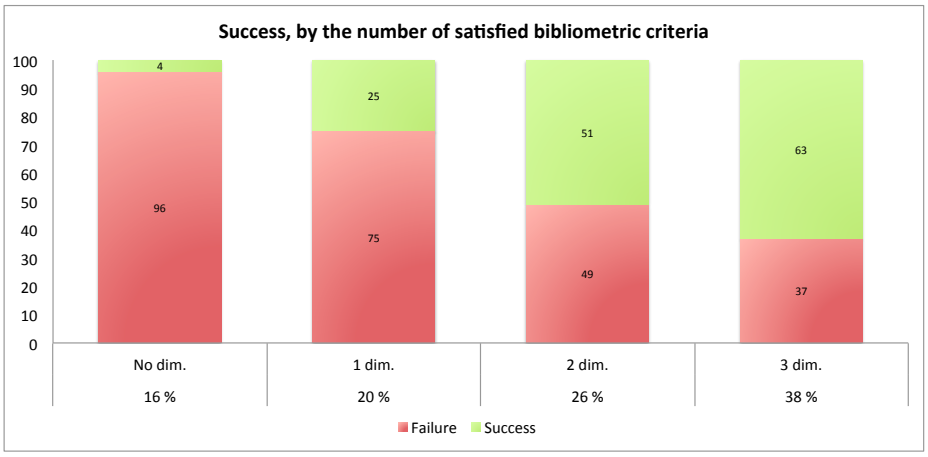
Zinovyeva, Natalia and Manuel Bagues (2014), "The Role of Connections in Academic Promotions," *American Economic Journals: Applied Economics*, forthcoming.

## Figure 1: Sample Individual Evaluation

**DOE John**

The candidate MARCO ROSSI has been Ricercatore universitario at the Università di XXX since 2006. His scientific work is concerned with the development of democracy, including a monograph on the role of public opinion in political thought and a series of contributions concerning English and Anglo-American thought and developments from the 17th through 19th centuries, with special reference to Edmund Burke. The candidate is a member of the "Re-Imagining Democracy in the Mediterranean, 1750-1860" project, based at the University of Oxford. The candidate has a significant number of international conference participations, among which those in which the English have invited him to speak about Burke are perhaps the most indicative of a strong international reputation. In terms of specific contributions, the "silent guest" metaphor is particularly significant in explaining how Burke plays out in the history of Italian political thought. The candidate scores above the median on two of the three indicators of impact and has substantial relevant teaching experience. On the basis of the application submitted, the candidate merits approval of the request for the abilitazione scientifica.

## Figure 2: Success rate and bibliometric measures




Note: Each bar provides information on the success rate of a group of candidates. Candidates are classified in four groups, depending on the number of dimensions where their productivity is above the median in the corresponding category. The first bar provides information for candidates who are above the median in every dimension. The second, third and fourth provide information for candidates that excel in one, two and three dimensions respectively.

**Table 1: Descriptive statistics – Evaluators**

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| | Mean | Std | Min | Max |
| | Eligible valuators | | | |
| Based in Italy: | | | | |
| - number | 32 | 19 | 8 | 128 |
| - share of women | 0.20 | 0.15 | 0 | 0.73 |
| Based abroad: | | | | |
| - number | 8 | 6 | 4 | 44 |
| - share of women | 0.12 | 0.15 | 0 | 0.75 |
| Expected share of women in committee | 0.18 | 0.14 | 0 | 0.69 |
| | Committee members | | | |
| Number of resignations per committee | 0.38 | 0.67 | 0 | 5 |
| Initial share of women in committee | 0.18 | 0.20 | 0 | 1 |
| Final share of women in committee | 0.19 | 0.20 | 0 | 1 |

Note: The table includes information from the 184 fields where qualification exams were held. The pool of *Eligible evaluators based abroad* refers only to 162 fields where the number of foreign based eligible evaluators was above 4.

## Table 2: Descriptive statistics – Candidates

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| | **Initial set of candidates (N=69,020)** | | | | | | |
| | All exams | | FP | AP | Male | Female | |
| | Mean | St.Dev. | Mean | Mean | Mean* | Mean* | p-value |
| **Individual characteristics:** | | | | | | | |
| Female | 0.38 | 0.49 | 0.31 | 0.41 | | | |
| Age | 44 | 8 | 49 | 43 | 0.01 | -0.02 | 0.000 |
| Based in Italy | 0.96 | 0.18 | 0.96 | 0.96 | 0.96 | 0.97 | 0.000 |
| Permanent position: | 0.55 | 0.50 | 0.74 | 0.47 | 0.55 | 0.56 | 0.190 |
| - same field | 0.75 | 0.43 | 0.77 | 0.74 | 0.74 | 0.78 | 0.000 |
| Relative order of application | 0.50 | 0.29 | 0.50 | 0.50 | 0.50 | 0.50 | 0.101 |
| **Quality indicators:** | | | | | | | |
| CV length, pages | 16 | 66 | 20 | 14 | -0.02 | 0.03 | 0.000 |
| All Publications: | 64 | 67 | 89 | 53 | 0.04 | -0.07 | 0.000 |
| - Articles | 36 | 51 | 52 | 29 | 0.07 | -0.11 | 0.000 |
| - Books | 2 | 3 | 2 | 1 | 0.05 | -0.08 | 0.000 |
| - Book chapters | 7 | 10 | 9 | 5 | 0.01 | -0.02 | 0.000 |
| - Patents | 0.24 | 1.65 | 0.35 | 0.19 | 0.02 | -0.04 | 0.000 |
| - Other | 20 | 32 | 25 | 17 | -0.00 | 0.00 | 0.822 |
| Average number of coauthors | 5 | 6 | 5 | 5 | -0.02 | 0.04 | 0.000 |
| Single-authored | 0.34 | 0.4 | 0.32 | 0.34 | 0.04 | -0.06 | 0.000 |
| First-authored | 0.22 | 0.2 | 0.22 | 0.22 | -0.00 | 0.00 | 0.793 |
| Last-authored | 0.12 | 0.16 | 0.15 | 0.11 | 0.02 | -0.03 | 0.000 |
| Total AIS (Sciences) | 61 | 90 | 86 | 49 | 0.05 | -0.09 | 0.000 |
| A-journal articles (SSH) | 4 | 8 | 7 | 4 | 0.04 | -0.05 | 0.000 |
| **Outcome variables:** | | | | | | | |
| Withdrawal | 0.14 | 0.35 | 0.16 | 0.13 | 0.13 | 0.17 | 0.000 |
| Failure | 0.49 | 0.50 | 0.48 | 0.50 | 0.49 | 0.49 | 0.413 |
| Qualified | 0.37 | 0.48 | 0.36 | 0.37 | 0.38 | 0.35 | 0.000 |
| | **Final set of candidates (N=59,150)** | | | | | | |
| **Production previous 10 years:** | | | | | | | |
| Social Sciences and Humanities | | | | | | | |
| - Articles | 20 | 19 | 25 | 18 | 0.04 | -0.06 | 0.000 |
| - A-journal articles | 3 | 4 | 3 | 2 | 0.04 | -0.05 | 0.000 |
| - Books | 2 | 4 | 3 | 2 | 0.07 | -0.10 | 0.000 |
| Sciences | | | | | | | |
| - Articles | 37 | 45 | 46 | 32 | 0.04 | -0.08 | 0.000 |
| - Citations | 60 | 102 | 77 | 52 | 0.02 | -0.04 | 0.000 |
| - H-index | 11 | 7 | 13 | 10 | 0.03 | -0.06 | 0.000 |
| Above median in all dimensions | 0.38 | 0.48 | 0.42 | 0.36 | 0.40 | 0.34 | 0.000 |
| Below median in all dimensions | 0.16 | 0.36 | 0.13 | 0.17 | 0.15 | 0.17 | 0.000 |
| **Evaluations:** | | | | | | | |
| Qualified | 0.43 | 0.49 | 0.43 | 0.43 | 0.44 | 0.41 | 0.000 |
| Unanimous decision | 0.80 | 0.40 | 0.79 | 0.80 | 0.81 | 0.78 | 0.000 |
| Individual votes | 0.44 | 0.47 | 0.46 | 0.44 | 0.45 | 0.43 | 0.000 |
| Length of collective evaluations | 280 | 231 | 304 | 269 | -0.01 | 0.02 | 0.000 |
| Length of individual evaluations | 176 | 277 | 203 | 164 | -0.00 | 0.00 | 0.119 |

Note: Detailed productivity indicators for the 10 years prior to exam were provided by the evaluation agency only for the sample of final applicants. These indicators differ for researchers in Science and in the Social Sciences and Humanities. Individual evaluations are not available for 3% of the candidates.
(*) In columns 5-6 productivity indicators and age are normalized at the exam level. Column 7 reports the p-value for the t-test of difference in means between genders.

**Table 3: Descriptive statistics – Share of women**

| | All | Permanent position in an Italian university: | |
|---|---|---|---|
| Sample of candidates: | | Yes | No |
| | | AP exams | |
| Potential | - | 0.45 | - |
| Candidates | 0.41 | 0.43 | 0.40 |
| Qualified | 0.39 | 0.40 | 0.37 |
| | | FP exams | |
| Potential | - | 0.35 | - |
| Candidates | 0.31 | 0.32 | 0.27 |
| Qualified | 0.29 | 0.30 | 0.22 |

Note: The set of potential candidates for Associate Professor and Full Professor positions includes respectively Associate Professors and Assistant Professors with a permanent position in an Italian university.

**Table 4: Performance of male and female candidates**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| | All | | | Category: | | Field: | |
| | | | | FP | AP | Sciences | SSH |
| Female | -0.022** | -0.004 | -0.006 | -0.002 | -0.007 | 0.001 | -0.014* |
| | (0.011) | (0.005) | (0.005) | (0.008) | (0.006) | (0.007) | (0.008) |
| Quality controls | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual controls | No | No | Yes | Yes | Yes | Yes | Yes |
| Exam FE | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Adj. R-squared | 0.001 | 0.374 | 0.395 | 0.386 | 0.400 | 0.426 | 0.344 |
| N | 59150 | 59150 | 59150 | 18061 | 41089 | 35856 | 23294 |

Note: OLS estimates. The dependent variable is an indicator for qualified candidates. Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, citations and H-index (in Sciences), A-journal articles and books (in Social Sciences and Humanities), dummies indicating whether candidates' observable productivity is above the median in the corresponding category and field, and the indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university. Standard errors clustered at the field level are reported in parentheses. *** denotes significance at 1%, ** significance at 5% and * significance at 1%.

## Table 5: Performance of male and female candidates - by evaluators' gender

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | All | FP | AP | Sciences | SSH |
| Female | 0.005 | 0.006 | 0.005 | 0.007 | 0.001 |
| | (0.007) | (0.011) | (0.008) | (0.009) | (0.013) |
| Female*Share of women in committee | -0.056** | -0.043 | -0.063** | -0.043 | -0.055 |
| | (0.028) | (0.046) | (0.028) | (0.032) | (0.045) |
| Quality controls | Yes | Yes | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes | Yes | Yes |
| Exam FE | Yes | Yes | Yes | Yes | Yes |
| Adj. R-squared | 0.396 | 0.386 | 0.400 | 0.426 | 0.345 |
| N | 59150 | 18061 | 41089 | 35856 | 23294 |

Note: OLS estimates. Standard errors clustered at the field level are reported in parentheses.
*** denotes significance at 1%, ** significance at 5% and * significance at 1%.

## Table 6: Randomization check

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Dependent variable: | CV length | Publ. | A-journal articles | Total AIS | Coauthors | Age | Appl. order | In Italy, same field | In Italy, other field | Based abroad |
| Female*Share of women in committee | 0.073 | -0.035 | -0.092 | 0.082 | -0.037 | 1.026* | -0.034 | 0.051 | 0.018 | -0.026 |
| | (0.089) | (0.085) | (0.104) | (0.115) | (0.070) | (0.589) | (0.023) | (0.038) | (0.021) | (0.017) |
| Female* Expected share of women | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Exam FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 69020 | 69020 | 27625 | 41395 | 69020 | 69020 | 69020 | 69020 | 69020 | 69020 |

Note: OLS estimates. The *Expected female share* among evaluators was computed using 1000 simulated draws from the pool of eligible evaluators, taking into account the constraint that committees cannot include more than one member from the same university. Standard errors clustered at the field level are reported in parentheses.
*** denotes significance at 1%, ** significance at 5% and * significance at 1%.

## Table 7: Does the gender of evaluators affect candidates' chances of success?

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Sample: | All | | | Category: | | Field: | |
| | | | | FP | AP | Sciences | SSH |
| | | | IV | IV | IV | IV | IV |
| Female | -0.015* | 0.000 | 0.005 | -0.001 | 0.008 | -0.000 | 0.010 |
| | (0.009) | (0.007) | (0.008) | (0.012) | (0.010) | (0.008) | (0.018) |
| Female*Share of women in committee | -0.073** | -0.100*** | -0.123*** | -0.101 | -0.134*** | -0.113*** | -0.125** |
| | (0.036) | (0.029) | (0.038) | (0.063) | (0.040) | (0.038) | (0.063) |
| Quality controls | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual controls | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Female*Expected share of women | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Exam FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adj. R-squared | 0.081 | 0.277 | 0.277 | 0.280 | 0.275 | 0.268 | 0.287 |
| N | 69020 | 69020 | 69020 | 21594 | 47426 | 41395 | 27625 |

Note: The dependent variable is an indicator for candidates that qualified. Columns 1 and 2 provide OLS estimates. In columns 3-7 the share of female evaluators in the final committee has been instrumented using the outcome of the initial random draw. Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, A-journal articles, and the indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university. Expected share of women in the committee is computed for an average of 1000 simulated draws from the pool of eligible evaluators, taking into account the constraint that committees cannot include more than one member from the same university. It is centered to have zero mean in the corresponding sample of candidates. Standard errors clustered at the field level are reported in parentheses. *** denotes significance at 1%, ** significance at 5% and * significance at 1%.

## Table 8: Do women shy away from male evaluators?

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| | | All | | Category: | | Field: | |
| | | | | FP | AP | Sciences | SSH |
| | | | IV | IV | IV | IV | IV |
| Female | 0.045*** | 0.028*** | 0.028*** | 0.036*** | 0.024*** | 0.028*** | 0.027*** |
| | (0.005) | (0.005) | (0.006) | (0.009) | (0.007) | (0.006) | (0.013) |
| Female*Share of women in committee | 0.012 | 0.019 | 0.023 | 0.024 | 0.022 | 0.023 | 0.013 |
| | (0.021) | (0.021) | (0.026) | (0.050) | (0.028) | (0.028) | (0.042) |
| Quality controls | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual controls | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Female*Expected share of women | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Exam FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Adj. R-squared | 0.046 | 0.109 | 0.109 | 0.140 | 0.091 | 0.098 | 0.125 |
| N | 69020 | 69020 | 69020 | 21594 | 47426 | 41395 | 27625 |

Note: The dependent variable is an indicator for candidates who withdrew their application before the evaluation took place. Columns 1 and 2 provide OLS estimates. In columns 3-7 the share of female evaluators in the final committee has been instrumented using the outcome of the initial random draw. Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, A-journal articles, and the indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university. Expected share of women in the committee is computed for an average of 1000 simulated draws from the pool of eligible evaluators, taking into account the constraint that committees cannot include more than one member from the same university. It is centered to have zero mean in the corresponding sample of candidates. Standard errors clustered at the field level are reported in parentheses. *** denotes significance at 1%, ** significance at 5% and * significance at 1%.

## Table 9: Evaluators' individual votes

| | 1 | 2 | 3 |
|---|---|---|---|
| Female candidate | -0.007 | 0.002 | |
| | (0.005) | (0.007) | |
| Female candidate*Share of women in committee | | -0.047* | |
| | | (0.025) | |
| Female candidate*Female evaluator | | | 0.006 |
| | | | (0.006) |
| Quality controls | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes |
| Exam FE | Yes | Yes | Yes |
| Application FE | | | Yes |
| Adj. R-squared | 0.407 | 0.407 | 0.897 |
| N | 274256 | 274256 | 274256 |

Note: Each observation corresponds to a candidate-evaluator pair. The dependent variable takes value one (zero) if the vote was positive (negative). Standard errors are clustered at the field level. Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, A-journal articles, and the indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university.

**Table 10: Length of Evaluation Reports**

| | 1 | 2 | 3 |
|---|---|---|---|
| | Length of ... | | |
| Dependent variable: | collective report | individual report | |
| Female candidate | 0.007* | 0.004 | |
| | (0.004) | (0.004) | |
| Female candidate*Share women in committee | -0.006 | -0.013 | |
| | (0.012) | (0.012) | |
| Female candidate*Female evaluator | | | 0.011 |
| | | | (0.020) |
| Quality controls | Yes | Yes | Yes |
| Individual controls | Yes | Yes | Yes |
| Exam FE | Yes | Yes | Yes |
| Candidate FE | | | Yes |
| Adj. R-squared | 0.847 | 0.881 | 0.726 |
| N | 58691 | 294738 | 294738 |

Note: The left-hand side is the log number of words contained in the collective report (column 1) or in the individual evaluation report of each committee member (columns 2 and 3). Quality controls include normalized CV length, publications (by type), average number of coauthors, proportion of first and last-authored articles, total Article Influence Score, A-journal articles, and an indicator for a permanent position in an Italian university in the same field. Individual controls include quadratic control for age, normalized application order, and indicators for researchers based abroad, exact position and university.