

Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment*

Manuel F. Bagues[†]
Universidad Carlos III

Berta Esteve-Volart[‡]
York University

October 2008

Abstract

This paper studies whether the gender composition of recruiting committees matters. We make use of the unique evidence provided by Spanish public examinations, where the allocation of candidates to evaluating committees is random. We analyze how the chances of success of 150,000 female and male candidates for positions in the four main Corps of the Spanish Judiciary from 1987 to 2007 were affected by the gender of their evaluators. We find that a female (male) candidate is significantly less likely to be hired whenever she (he) is randomly assigned to a committee where the share of female (male) evaluators is relatively greater. Evidence from multiple choice tests suggests that this is due to the fact that female majority committees overestimate the quality of male candidates.

Keywords: gender discrimination, randomized experiment, public examinations.

JEL Classification: J71, J78.

*We are particularly grateful for the useful comments and suggestions provided by Samuel Berlinski, Sonia Bhalotra, Tim Besley, Alison Booth, Juan José Dolado, Florentino Felgueroso, Nicole Fortin, Lawrence Katz, Manuel Miranda Estrampes, Nathan Nunn, John Palmer, María José Pérez Villadóniga, Aloysius Siow and all of the participants in our seminars and conference presentations at Universidad Carlos III, the University of Toronto, the University of British Columbia, the University of Victoria, Université Louis Pasteur, York University, WZB Berlin, Erasmus University, SOLE Chicago, EALE Prague, ZEW Mannheim, SAMES Chennai, EEA Budapest, and SAE Asturias. We also thank the judges who attended the May 2006 seminar at the *Centre d'Estudis Jurídics i Formació Especialitzada* in Barcelona and the July 2006 seminar at the *Escuela de Verano del Consejo General del Poder Judicial* in A Coruña. The second author thanks the University of British Columbia and Universitat Pompeu Fabra for all of their hospitality.

[†]*mfagues@emp.uc3m.es*. Universidad Carlos III, Departamento de Economía de la Empresa, Calle Madrid 126, 28903 Getafe (Madrid), Spain.

[‡]*berta@econ.yorku.ca*. York University, Department of Economics, 4700 Keele Street, Toronto M3J 1P3, Ontario, Canada.

1 Introduction

Legislation encouraging gender quotas in top positions has been adopted in some countries and is being considered in many others. In Norway, publicly appointed committees, boards, and councils have been required since 1988 to be made up of at least 40% of each gender. This requirement was extended to all shareholder-owned companies' boards of directors in January 2008. The French Parliament passed legislation in 2001 mandating gender parity in party lists for a variety of elections. In Spain, newly elected Prime Minister José Luis Rodríguez Zapatero appointed women to half of his cabinet posts in 2004. Furthermore, in March 2007 the Spanish Equality Law was passed, imposing gender parity in all selection committees in the state administration, party lists, public organizations and related firms. Private corporations in Spain also received governmental guidelines on how to achieve greater participation of women on boards.¹ In Chile, in 2006 newly elected President Michelle Bachelet also appointed women to half of her cabinet posts.

The reasons for imposing gender parity in top positions lie in the extremely low percentage of decision-makers who are women, within both the public and private spheres. In politics, only in 19 out of 189 countries did women account for 30% or more parliamentary seats in 2007.² In Italy and France, respectively, only 3% and 4% of the 50 largest companies' board directors are women.³ In the US, women made up only 3.4% of the top-level management in 1997 (Bertrand and Hallock 1999).

In the past, policy towards gender equality was focused on the so-called *equal opportunities* approach.⁴ Underlying this approach was the pipeline theory, according to which women must move their way through a metaphorical pipeline to reach top-level jobs. Accordingly, policy was designed to encourage women's higher education on the understanding that providing women with the same human capital as men would enable them to reach the top positions they seemed unable to attain. Evidence supporting the pipeline theory, however, is disappointing. For instance, in all the economics PhD-granting departments in the United States, the incidence of women among new PhDs was 24.1% in 1996 and rose up to 31% in 2006, while women accounted for only 8.4% of the country's full economics professors in 1996 and even less of them, 8.3%, in 2006.⁵ In the same vein, there is a prevailing view that while women have started to move up into management and public service positions, once they reach a certain point, the so-called 'glass ceiling', they do not seem to go any further.

Pessimism about the pipeline theory might explain the more recent approach: the imposition of gender quotas in top positions.⁶ How might gender quotas work? A gender quota at the top level automatically equalizes the numbers of men and women in top positions. This helps close the gender gap quite directly, but a further motivation for imposing gender parity in top positions only (as opposed to all positions) is the rationale that once more women fill those positions, it should be easier

¹Official State Bulletin (BOE) 71, March 23, 2007, p.12611.

²Inter-Parliamentary Union (February 2007), <http://www.ipu.org/wmn-e/arc/classif280207.htm>.

³*The Economist*, November 25, 2005, citing a report by the Aspend Institute Italia.

⁴In the US and other countries affirmative action policy involving quotas was also implemented, but generally not at the 50-50 level (Fryer and Loury 2005).

⁵*Report of the Committee on the Status of Women in the Economics Profession* (CSWEP Newsletter, Winter 2006).

⁶The shift in policy is obvious in the Spanish case. In a recent governmental document on proposals towards gender equality, equality of opportunity is mentioned only once, while gender parity is mentioned six times (ORDEN PRE/525/2005, BOE, March 8, 2005).

for other women to advance through the lower ranks and ultimately reach the top themselves. That is, gender parity in decision-making positions could break the ‘glass ceiling’ from above, having a knock-on effect for women working their way through lower echelon jobs.

There are several ways in which this could happen. First, women at the top could become role models. If women are not currently getting to top positions because of social norms, having more women at the top might help change these social norms. Second, women in top positions can affect choices in ways that might help other women get to the top. For instance, they could choose more flexible working hours or promote public expenditure that benefits women more (in line with evidence in Chattopadhyay and Duflo 2004). More directly, women who get to top-level positions because of gender quotas might hire more women than their male counterparts. This will be the case if female candidates are more likely to be recruited when evaluated by female evaluators. It is this latter proposition that is the focus of the present paper. Although implicit in many discussions of gender parity policy,⁷ there is no clear evidence supporting the hypothesis that female evaluators are more favorable towards female candidates.

A neat empirical analysis of the effects of gender parity is hard to come by. In most situations, the composition of hiring committees is not casual and might be potentially related to either the position or candidates’ characteristics. For this reason, it is not usually possible to establish causality: is the committee composition affecting the hiring gender balance, or does there exist some unobservable factor which determines both the choice of the committee and the gender balance of the hiring? In order to avoid such endogeneity problems, here we take advantage of the random assignment of candidates to evaluation committees used in public examinations in Spain. These exams, also common in other countries in continental Europe, Asia and Latin America, typically involve an extremely large number of candidates. In general, several evaluation committees have to be formed, and a lottery determines the allocation of candidates to committees.

In this paper we use information on the outcome of 51 public exams used to make appointments to four different Corps of the Spanish Judiciary from 1987 through 2007, involving 2,467 evaluators and approximately 150,000 candidates. In addition to the fact that the existence of a mechanism of random allocation eliminates endogeneity concerns, this study also benefits from the fact that the subjects and the experiment are actually taken from real life, with the subjects receiving very substantial payoffs if they obtain the positions for which they are applying. Hence we are able to avoid the usual problems associated with artificial settings. Moreover, the repeated nature of the experiment—over two decades and across different types of positions—also allows us to test whether the effect of the

⁷A recent report of the Spanish Securities and Exchange Commission, supporting the introduction of a gender parity policy, argues that the low percentage of women on boards is due to the existence of ‘old boys’ network effects and hysteresis (“el predominio de hombres en puestos de responsabilidad es un fenómeno en parte auto-inducido, sujeto potencialmente a histéresis y externalidades de red esto es, el predominio de hombres (*old boys networks*) hace más probable que se sigan nombrando consejeros de ese sexo, de forma que la falta de diversidad no se corregirá sin un esfuerzo deliberado para lograrlo” (*Proyecto de Código Unificado de Recomendaciones de Buen Gobierno de Sociedades Cotizadas, Comisión Nacional del Mercado de Valores*, January 2006).

Similarly, another governmental publication argues that the lack of females in Spanish academia is due to male evaluators discriminating against female candidates (“Tanto en el CSIC como en la Universidad, la elección de sus miembros se basa en realidad en un sistema de cooptación, disfrazado de concurso de méritos.... Este sistema parece beneficiar más a los hombres que a las mujeres, pues las barreras surgen en el momento que otros, fundamentalmente hombres, juzgan la idoneidad y niegan la entrada a las mujeres en las categorías más altas”, *Mujer y Ciencia: La situación de las Mujeres Investigadoras en el Sistema Español de Ciencia y Tecnología*, FECYT, p. 48).

gender composition of committees has changed over time or whether it varies according to the degree of feminization of the position.

Our main finding is that a female (male) candidate is significantly less likely to be hired whenever she (he) is randomly assigned to a committee where the share of female (male) evaluators is relatively greater. This result is in line with previous work by Broder (1993), who finds that female authors applying for grants to the US National Science Foundation have lower chances of success when evaluated by female reviewers than when evaluated by their male colleagues. In our case, and unlike Broder (1993), the evaluations received by male candidates tend to be higher when there are relatively more women in the committee. This, in turn, reduces the chances of success of female candidates applying for the same positions.

Evidence from multiple choice tests suggests that our results are consistent with two hypotheses: (1) female evaluators tend to overestimate the quality of male candidates; (2) the presence of women in committees affects the voting behavior of their male colleagues such that male members increasingly favor male candidates. Unfortunately, given that we can only observe the final decision of committees but not the individual voting behavior of committee members, we cannot disentangle these two hypotheses. We also find that the bias we observe is stable over time, and that it does not depend on the degree of feminization of positions.

The empirical evidence presented in this paper is relevant for a number of reasons. According to our study, candidates' probability of success may not necessarily be higher when evaluated by evaluators of their own gender, as may be commonly thought. While the data are based on recruitment of civil servants in Spain, it is possible that other hiring processes elsewhere would show similar patterns.

The Spanish government has recently decided to impose gender parity in all public recruiting committees, including the committees we study here. This policy is aimed at increasing the number of successful female candidates.⁸ However, our results suggest that gender parity in recruitment committees will not increase the incidence of women in top positions. In fact, our calculations for the 1987-2007 period show that, had there been an additional woman in every committee for the exams we study here, around 123 (2.8%) fewer women would have been hired.

Our 'experiment' constitutes an example of a randomized design that operates in an important framework: the elite formation of society's elite.⁹ Successful candidates in the public examinations considered here make key decisions affecting society. This is particularly important in countries like Spain, where public positions are for life.

The paper is organized as follows. Section II describes the related literature. Section III offers background information on public examinations in Spain, and Section IV describes the data. Section

⁸“(…) The so-called, in the Law, principle of balanced presence or composition, which is aimed at reaching a significant representation of each gender in top positions” [(…) El llamado en la Ley principio de presencia o composición equilibrada, con el que se trata de asegurar una representación suficientemente significativa de ambos sexos en órganos y cargos de responsabilidad.] (Equality Law, BOE, 71, March 23, 2007, p.12611.)

⁹Furthermore, many political figures in Spain served in the Judiciary before landing important posts in the government. This is particularly remarkable given recent literature showing that the composition of the government is not irrelevant to policy choices, and thus suggesting that women might have different preferences on public expenditure than men (Pande 2003, Chattopadhyay and Duflo 2004).

V turns to the empirical analysis and Section VI discusses results. Finally, Section VII concludes.

2 Related literature

There exists a large body of literature providing empirical evidence which indicates that the gender of candidates might matter. Blank (1991) compares single-blind and double-blind reviewing of papers submitted to *The American Economic Review* and finds a small, insignificant effect, in that female authors fare better under double-blind reviewing. Ayres and Siegelman (1995) organize an experiment in which test buyers are randomly allocated to new-car dealerships to negotiate over the price of a car and find that female and black test buyers are discriminated against by the dealers. Goldin and Rouse (2000) find that the adoption of a screen in orchestras' hiring of musicians fosters impartiality in hiring and increases the proportion of women hired.

There also exists a smaller, but growing, literature in economics that studies whether various characteristics of evaluators matter.¹⁰ In the contribution most closely related to ours, Broder (1993) examines the ratings of economics proposals to National Science Foundation grants in the US by gender of the applicant and gender of the reviewer. She finds that female reviewers rate female-authored NSF proposals lower than do their male colleagues. Dillingham *et al.* (1994) find that, in the context of elections held by a professional organization of economists, female voters are more likely to vote for female candidates. Lavy (2008) compares data on blind and non-blind scores that high school students receive on matriculation exams in their senior year in Israel and finds that the grades obtained in non-blind tests are sensitive to the characteristics of evaluators. Antonovics *et al.* (2005) study voting behavior in a television game show and find that women are less likely to vote against females in the early rounds of the game.¹¹ Price and Wolfers (2007) investigate whether the race of NBA referees affects how they treat black and white players, and find that more personal fouls are called against players when they are officiated by an opposite-race refereeing crew than when officiated by an own-race crew.

The link between discrimination and the evaluator's gender has also received attention from other fields. Social psychologists draw on two main motives for such discrimination.¹² The similarity-attraction paradigm (Byrne 1971) posits that individuals who are similar will be interpersonally attracted. Thus, female evaluators might then be more *attracted* to female candidates. On the contrary, the *self-enhancement drive* motive posits that members of lower status groups may seek to identify with the higher-status group. Thus, in male dominated fields female evaluators might identify with males to maintain a positive social identity (Ely 1995) and, accordingly, would then be more favorable to male candidates (Graves and Powell 1995).

¹⁰Unfortunately some of the best known papers on gender discrimination in the field of economics do not deal with the interaction between the characteristics of evaluators and those of candidates. According to private correspondence with the authors, in the case of Blank (1991) this was not possible due to the small size of the sample of female referees reviewing female-authored papers. In the case of Goldin and Rouse (2000), the lack of information on the juries made it impossible to test whether the observed gender bias was affected by the gender composition of committees.

¹¹See also Heilman *et al.* (1988) and Wiley and Eskilson (1985).

¹²See Goldberg (2005) for a detailed explanation and review of these motives.

The empirical evidence on whether the evaluators' gender matters for candidates' success is, however, often contradictory and usually based on small, localized samples. First, some authors fail to find any effect. Gallois *et al.* (1992) find that personnel officers who viewed videotapes of mock interviews perceived same-sex candidates to be more similar to themselves (consistent with the similarity-attraction motive), but this did not affect the evaluators' ratings of suitability for the job. Similarly, in an experimental evaluation of fake applications for a assistant principal position, Bon Reis *et al.* (1999) do not find any significant difference in the evaluations performed by female and male high school principals. Second, some studies find that sex similarity is positively related to selection decisions. Graves and Powell (1996) analyze data from 680 actual interviews by recruiters at a large American university and find that female recruiters evaluate female applicants more favorably. Finally, some studies provide evidence consistent with the self-enhancement drive motive: Graves and Powell (1995) and Goldberg (2005) examine the effect of sex similarity on recruiters' evaluations of actual applicants and find that female recruiters tend to perceive male applicants as being more similar to themselves and more qualified than female applicants. Likewise, Ely (1994) uses a small sample of law firms and finds that, compared with women in firms with many senior women, women in firms with few senior women were less likely to experience common gender as a positive basis for identification with women.¹³

The evidence described above generally suffers from at least one of two problems. While in most observational studies there are obvious endogeneity problems, sometimes experimental studies may not be suitable to study labor market decisions (Palacios-Huerta and Volij 2008). In this respect, a valuable setting would be one taken from the real labor market, yet also using an experimental design. Moreover, as suggested by the social psychology theories, the effect of the evaluator's gender on her evaluations may vary depending on the context: women with a more masculine attitude towards work are more likely to enter, and remain in, masculine occupations. Consequently, female recruiters in such occupations may identify more with men than with other women. Hence, one would want to test whether an evaluator's gender matters over a variety of situations with varying degrees of female abundance.

Public examinations in Spain provide unique evidence in both respects. Candidates are allocated to committees through a random lottery, overcoming any concerns about the endogeneity of the treatment, and the result of the evaluation has relevant implications. Furthermore, our database includes practically all public exams held in the last twenty years to the main positions of the Spanish Judiciary, allowing us to observe not one single case but a number of "experiments" which are repeated over time and across fields. This makes it possible to test, for instance, whether evaluators' bias is affected by the relative *masculinity* or *femininity* of the field.

¹³Another related strand of literature is that of student achievement—indeed, while this literature does not focus on evaluation bias, it does consider similarity-attraction to explain differences in outcomes such as student scores (Dee 2004).

3 Background

Nation-wide public exams have traditionally been used as the method of evaluation to determine access to a variety of public positions in many countries in continental Europe, Asia and Latin America. In Spain, obtaining a permanent position in the public sector requires success in the corresponding public examination. In total, approximately 250,000 individuals participate every year in public examinations in Spain.¹⁴ Here we use data from public examinations for four of the most coveted positions in the Spanish Judiciary: the notary, judge, prosecutor and court secretary positions. These exams, which are held every one or two years, are taken by large numbers of candidates, the probability of success is low, typically about 5% per year (Bagues 2005), and failing candidates tend to retake the exam. These exams are organized at the national level and are typically held in the capital, Madrid.

3.1 Women in the Spanish Judiciary

The Spanish Judiciary's top positions provide us with a very interesting setting in which to study gender issues. During most of the Franco regime and until 1964, women were banned by law from holding positions in the Public Administration.¹⁵ After the ban was lifted it still took a few years for the first women to enter the Spanish Judiciary,¹⁶ and it was not until the advent of democracy in the late 1970s that females irrupted into the Judiciary's ranks in significant numbers (see Figure 1). By the mid 1980s some female candidates already ranked first in several entry examinations,¹⁷ and by the 1990s the majority of successful candidates in most exams were female. This increase in the incidence of women in the Judiciary has been parallel to the increase in the number of women graduating from law schools. As Figure 2 shows, in 1973 less than 20% of Spain's law graduates were women. In 2004 this percentage was above 60.

The evolution of the incidence of women in judicial positions in the last four decades is thus consistent with a crack in the glass ceiling, at least at the entry level. However, progress has not been easy. The pioneers tended to encounter opposition as gender prejudice still persisted in some layers of the profession. In 1972, the president of the Superior Court of Justice declared that women were unsuited to the task of adjudication: "(judging) is in essence a male task that can collide with female sensitivity".¹⁸ (Even now, the dictionary published by the *Real Academia de la Lengua Española* still reports the female word for judge in the Spanish language, *jueza*, as a colloquial term for the [male] judge's wife.)¹⁹ The first female candidate to rank number one in the exam for judge positions recounts how in 1981 Pérez de Petinto, a reputed forensic surgeon who used to award the top candidate with a book, awarded her with a book titled *Spanish Women*—inside, Petinto wrote that the candidate's main

¹⁴Ministry of Public Administration, http://www.map.es/prensa/notas_de_prensa/notas/2007/02/2007_02_02_2 (retrieved October 2, 2008).

¹⁵*Ley de Funcionarios Civiles del Estado*, Decreto 315/1964.

¹⁶According to the 1995 corresponding Rankings of Public Servants (*Escalafón*), females have been present in the court secretary Corps since at least 1967, when Maria Dolores Mosqueira Riera entered. In 1974 the Asturian Belén del Valle became the first female to pass the entry exam to judge and prosecutor positions.

¹⁷In 1980 Concepción Jarava Melgarejo became the first woman to rank first in the notary entry exam; so did Margarita Robles in 1981 in the judge and prosecutor exams.

¹⁸*La Revista*, El Mundo, 177, 1999.

¹⁹RAE, Diccionario, 22nd edition.

ambition should always be to become a ‘good wife and mother’. The (male) candidate who ranked second was awarded the Quixote—for his chivalry in *giving* a female the first position.²⁰ Nevertheless, survey information suggests that most members of the Judiciary welcomed the arrival of women: in a survey of judges, 72% in 1984 and 73% in 1990 answered the question *do you think that the increase in female judges is right?* with a positive answer; only 4% and 2% of them, respectively, answered negatively.

Finally, note that while the incidence of women has grown over time in each of the four fields considered, patterns have differed by Corps: notary positions constitute a male dominated field even today, judge and prosecutor positions are mixed, and court secretary positions are disproportionately filled by women.

3.2 Structure of Public Exams

To understand how the gender composition of committees can have an effect on candidates’ success, we must first explain how public examinations function. Below we explain (i) how candidates are allocated to evaluation committees, (ii) how committees are formed, (iii) how the evaluation process is structured, and (iv) how final grades are set.

In the remainder of the paper, we denote by “exam” every public examination process from the beginning to the end, and we denote by “test” every stage within a given exam (e.g., in the year 2004 one judge and prosecutor exam was held, consisting of three tests; one multiple choice and two oral).

3.2.1 Allocating candidates to committees

Applicants to any top Corps of the Judiciary must have an undergraduate degree in law. The large number of applicants usually requires the creation of multiple evaluation committees. Since 1987, an evaluation committee has been created for approximately every 500 candidates. Once committees have been formed, every committee is assigned a given number of candidates and positions. In every case the number of candidates assigned to each committee is proportional to the number of positions initially assigned to the committee, which constitutes the maximum number of positions that each committee can initially allocate. All committees within a given exam are generally of equal size; however, due to indivisibilities, the number of candidates in each committee is not necessarily the same; it can vary slightly.

The allocation of candidates to committees always follows some random procedure. For the notary exam, the allocation is decided by a random lottery which directly matches candidates to committees. In the exams for judge, prosecutor and court secretary positions, candidates are ranked in alphabetical order and committees are ranked numerically. A lottery then decides the initial according to which the alphabetical list of candidates will be matched with the list of evaluation committees.²¹ For instance,

²⁰ *La Revista*, El Mundo, 177, 1999.

²¹ Given the nature of this matching process, it is possible that there exist differences across groups due to the

in the exam for judge positions held in the year 2000, the randomly chosen letter was “B”; hence, the first candidate in the list whose initial was “B” was matched to the first committee in the list, and so on.²²

3.2.2 Committee composition

Table 1 displays information on the rules of committee composition and the format of the examination, by position. Committees are formed by between seven and ten members. The number of members is the same for every committee within a given exam but can vary slightly over time or across positions. Members are appointed according to rules specifying their Corps of origin and their qualification. Each committee is composed of, first, members of the Corps that offers the position in question; second, members of some other Corps of the Judiciary (notaries, registrars, judges, prosecutors, court secretaries, and public defenders); finally, external members (law professors and private sector lawyers). Note that members of the Corps that organizes the exam never make up more than half of the committee.²³

In general, each Corps of origin selects its evaluators among those members of the Corps who are eligible and have volunteered for the task. Thus, even though the committee composition itself is not determined by a lottery among potential evaluators, each Corps assigns members to exams in an independent manner.

For a variety of reasons, approximately 10% of the evaluators originally appointed to committees tend to be replaced before the evaluation process starts. In most cases individuals initially appointed as committee members quit due to promotions or other appointments. Likewise, committee members must resign if a candidate with whom they share close family ties is assigned to their committee.

3.2.3 Format and structure of the exam

The structure of the process is very similar in the four public examinations analyzed here.²⁴ Exams are composed of several qualifying stages; in each stage candidates are evaluated on a set of topics. The list of possible topics is very long, usually close to two hundred. Indeed, preparing public examinations requires learning thousands of pages of law articles by heart. Candidates are expected to memorize the law articles and then *regurgitate* them during the examination. A random lottery decides which topics a candidate must answer. The lottery consists of numbered balls corresponding to the topics in the test. Five balls are typically drawn, determining a particular five-question test. Candidates must answer all five questions within around one hour in total. Once this hour has elapsed, members of the committee can potentially ask for clarification from the candidate, however such interactions between

sorting on alphabetical grounds. However, this is orthogonal to gender: there is no reason to expect that differences in family-name initials are correlated with gender.

²²BOE, 196, 16/8/2000, p. 29164.

²³*Ley Orgánica 6/1985*, July 1st.

²⁴More precise details about the structure of each exam are provided in the appendix.

the candidate and the committee are in practice very rare.²⁵

During our period of study there have been some changes in the way exams are carried out (see Table 1). Until 1995, in the exams for judge, prosecutor and court secretary positions, candidates were first given several hours to write their answers, and were then required to read them in front of the committee. During the reading stage, a public clerk kept a copy of the original test to make sure that the candidate would not change the written version while reading it.²⁶ After 1995, candidates taking these exams were required to give oral answers directly, without first writing them.²⁷ Moreover, since 2003 the exam for the judge and prosecutor positions includes a preliminary multiple choice test. The same is true for the exam to court secretary positions since 2006. The structure of the notary exam has remained the same over the period of study. The process is composed of two oral and two written tests; the latter must again be read by candidates in front of the committee.

3.2.4 Grading

In each stage a candidate receives an evaluation if she manages to answer all questions—something that many candidates fail to do. In the case of candidates who manage to answer all questions, the committee decides on a majority basis whether the candidate has passed; in case of a tie, the committee president rules. In addition to the pass or fail decision, a numerical grade is also assigned to candidates who pass the test. This numerical grade is calculated through a voting process, with each committee member casting a ballot with a proposed grade. For each candidate, the minimum and the maximum grade ballots are excluded, and the final grade is calculated as the average of the remaining ballots. Even though committee members vote on an individual basis, committee members may discuss their decisions prior to making them.

Passing all stages of the public exam is a necessary condition to obtaining a position but it is not always sufficient. All passing candidates will receive positions only when there are a sufficient number of available positions to accommodate them. When there are more passing students than available positions, the selection is based on final rankings calculated using the candidates' final grades. A candidate's final grade is the sum of grades that she has obtained in each stage of the exam.²⁸ Once final grades are calculated, these grades are used to rank each successful candidate within his or her committee. Grades are also used to establish an overall ranking of all candidates: candidates ranked in first place in each committee are ranked in relation to each other, based on their final scores. Next, candidates ranked in second place in each committee are then ranked in relation to each other, again based on their final scores. This process continues until all successful candidates have been ranked.

²⁵This information is based on personal conversations between the authors and a number of evaluators and candidates. Note that this is not the case in other entry exams to the public administration. In this respect see the work by Quintero (2008).

²⁶BOE, 31, 5/2/1987, pp. 3525-31.

²⁷BOE, 166, 13/7/1995, pp. 21545-630.

²⁸For the judge and prosecutor exam and the court secretary exam, a (qualifying) preliminary stage was introduced in 2003 and 2006 respectively consisting of a multiple choice test. The grade obtained in the multiple choice test does not count towards the final grade.

4 Data

Our database contains information on practically every exam held between 1987 and 2007 in the four main Corps of the Spanish Judiciary: judge, prosecutor, notary and court secretary.²⁹ Exams are held every one or two years. The judge and prosecutor exams were separate until the year 2001. Since then there has been a unique exam covering both judge and prosecutor positions. Exams held before 1987 have been excluded from our database since typically there was only one committee per exam at this time, and that prevents us from exploiting the random allocation of candidates to committees.³⁰ Below we describe the available data in detail.

4.1 Exams

Table 2 shows information on the public examinations for which we have data. Our database includes 51 exams. The average exam in our database consisted of six committees, and in total we have information on 309 committees. Since a committee is typically created for approximately every 500 candidates, about 150,000 candidates were evaluated in these exams (an average of 3,000 candidates per exam). The size of the exams varies across fields: notary exams are relatively small, with an average of two or three committees per exam, while judge exams are the largest, with an average of around ten committees per exam. Exams for the prosecutor position and joint exams for the judge and prosecutor positions were composed of, on average, seven committees. Finally, court secretary exams usually consist of about four committees per exam.

In 25 out of the 51 exams the number of candidates who passed every test was smaller than the number of available positions at the exam level, and hence every passing candidate was automatically assigned a position. In the remaining 26 exams the number of candidates passing every test exceeded the number of positions. In these cases, candidates were assigned to positions according to the ranking method discussed above.

Exams also differ in terms of format. In 20 of the exams included in our database, all tests were first written by candidates and then read to committees. In 21 exams, all tests were directly performed orally. Six of these exams also included an initial multiple choice test. The remaining ten exams, all of which were for notary positions, included two oral and two written tests (where, again, a written test involves writing it first and then reading it to the committee).

4.2 Candidates

For most of the period 1987-2007, information is only available for successful candidates—i.e., those candidates who are actually appointed to positions—and not for those candidates who failed. Multiple

²⁹Unfortunately, it was not possible to obtain information on the exams to court secretary positions held in 1991, 1993 and 1997.

³⁰Similarly, we have not considered exams for the other two high Corps of the Judiciary, the registrar and state lawyer Corps, because they seldom involve more than one committee.

choice tests were introduced in the court secretary exam (in 2006) and the judge and prosecutor exam (in 2003). We have gathered information for every candidate (successful or not) who took those exams. That is, we have information on every candidate who took the judge and prosecutor exam over 2003-2007, and every candidate who took the court secretary exam in 2006.

4.2.1 Successful candidates, years 1987-2007

Between 1987 and 2007 7,700 candidates turned out to be successful; this figure amounts to approximately 80% of the current members of these Corps. Table 3 shows the distribution of successful candidates across types of exams and provides descriptive statistics at the individual level on gender and the grades obtained by type of examination. Note that the information reported here refers only to those candidates who managed to obtain a position, so it should be interpreted bearing in mind that these candidates reflect only the upper part of the distribution (about 5% of all candidates). In total approximately 60% of candidates are female (the figure ranges between 38% of women in the notary exam, and 69.3% in the court secretary exam). Average grades of successful candidates are between 0.56 (judge candidates) and 0.71 (notary candidates).

Table 4 displays descriptive statistics at the committee level. It shows the number of positions initially available through each committee and the number of successful candidates, in total and by gender, for the exams included in our study. On average there were about 27 positions available per committee, 25 of which were ultimately assigned to some candidate.³¹ Among the positions assigned by each committee, an average of approximately 15 were assigned to women and ten to men.

Figure 3 shows the evolution over time of the fraction of successful candidates who were female by exam and committee. The incidence of female candidates has increased continuously. While in the late 1980s most of the successful candidates in the judge, prosecutor and notary exams were still male, by the 1990s the majority of successful candidates in the exams to judge and prosecutor positions were female. The notary exam has followed a similar pattern, although in this case female successful candidates only outnumbered male successful candidates in 2003. The graph also shows that the proportion of successful candidates who are female varies considerably across committees in a given exam. For instance, if we consider the ten committees that were formed for the judge exam in 1999, in the first committee 23 out of 31 successful candidates were female. In committee number eight there were only 13 females amongst the 28 successful candidates. In the remaining eight committees formed that year, the percentage of female successful candidates ranged between these two figures.

³¹The number of candidates per position is always the same across the different committees that evaluate an exam, but the precise number of candidates and positions that have been assigned to each committee can differ slightly because of indivisibilities. It should also be noted that in three of the examinations analyzed here one particular committee was assigned half as many positions and half as many candidates as the rest of the committees.

4.2.2 Candidates for judge and prosecutor (2003-2007) and court secretary (2006) positions

In Table 5 we display descriptive statistics on all individuals who registered for the exam for judge and prosecutor positions in the years 2003-2007, and those who registered for the exam for court secretary positions in 2006. In total there were 24,530 candidates, 30% of whom were male. Approximately 23% of the candidates were taking the exam for the first time. Similarly, 20% had already taken it (and failed) once or twice, and 47% the candidates had already taken the exam previously at least three times.

As explained above, both the exam for judge and prosecutor positions (since 2003) and the exam for court secretary positions (since 2006) are divided into three stages, all of which must be passed in order to qualify for a position. The first stage, a preliminary multiple choice test, was taken by 83.7% of the candidates who had originally registered for the exam. The minimum grade required to pass the multiple choice exam is fixed ex post based on the number of evaluation committees that are available for the second and third stages. On average about half the candidates taking the multiple choice test pass it. Both the second and third stages are oral. The threshold to pass each of these tests is of 12.51 points out of a maximum of 25 points. Out of all candidates registered, only about 10% managed to pass stage two. In total, only 5.1% of the candidates who registered for the exam passed the third stage. During the period we study, in every case except for one all candidates passing the third stage got a position.³² Male candidates tended to be slightly more successful (5.3% vs 4.9%).

4.3 Committees

Descriptive statistics regarding the characteristics of the 309 evaluation committees included in our sample are shown in the bottom panel of Table 4.³³ The average committee is formed by approximately eight members.³⁴ Of them, on average about 22% were women. We distinguish six different groups of committees according to their female share: committees in which (i) less than 10% of committee members are women (or, equivalently, committees with no woman); (ii) there are at least 10% of women, but less than 20% (committees with one woman); (iii) there are at least 20% of women but less than 30% (committees with two women); (iv) there are at least 30%, but less than 40% of women (committees with eight, nine or ten members, out of which three are women); (v) there are at least 40%, but females are not majority in the committee (committees with seven members, out of which three are women, and committees with eight or more members, out of which four are women); (vi) there is female majority, that is, over half of members are female.³⁵ According to this classification, the distribution of women in committees is the following: in 20% of committees there are no female

³²In the 2004 judge and prosecutor exam, there were 241 candidates who passed the three stages yet only 240 positions were available.

³³In what follows we use information about the composition of committees at the time when the evaluations took place. The picture would be almost identical if the initial (pre-evaluation) composition was used instead and committee member replacements were thus not considered.

³⁴The set of candidates and evaluators is not completely different from one year to another. Both (unsuccessful) candidates and evaluators may participate several times. The total number of unique evaluators observations is 1,301.

³⁵In a few cases where the percentage of women and men in the committee was exactly 50%, we have allocated them to group (v) or (vi) according to the gender of the president, whose vote prevails in case of tie.

members, in almost 30% of committees there is only one woman, and in 19% of cases there are two women. There are fewer committees in the other groups: in 13% of committees there is between 30% and 40% of women, and in 14% between 40% and 50%; finally, only in 5% of committees there is a female majority.

Figure 4 displays the number of female evaluators by year and type of examination. The number of female evaluators has increased over time in the four cases considered. Despite this increase, men still outnumber women in most committees; only in the court secretary exams the numbers of female and male evaluators are balanced. For the other exams, the average percentage of women is still between 20 and 30%. However, a comparison of Figures 1 and 4 suggests that the incidence of women as evaluators in public examinations does not differ significantly from the incidence of women in their respective Corps.

Figure 4 also reflects the fact that the number of female evaluators can differ greatly across committees within the same exam. For instance, candidates who took the judge and prosecutor exam in 2004 could have been assigned, depending on the lottery, to a committee with one, two, three, four or five female evaluators, out of a total of nine committee members.

In addition to gender composition, in Table 4 we also observe some other characteristics of individuals who served in committees. Namely, we observe whether they served in a similar evaluation committee within the previous three years and, for all evaluators who are members of the Judiciary—that is, all members of committees except for private lawyers and university professors—we observe their age and ranking.³⁶ On average, committee members have served in a similar committee 0.62 times in the previous three years. The average age of committee members is about 46 years. Given that age and ranking are highly correlated in our data, ranking has been redefined as the ranking of the evaluator relative to all other members of her Corps who were born the same year. This measure is normalized to be equal to one (zero) if the individual ranks first (last) among individuals born the same year. Committee members are relatively highly ranked (0.63) with respect to the population of all members in their Corps, whose average ranking is, by construction, equal to 0.5.

Table 6 presents descriptive statistics about these characteristics by gender and profession of the evaluator. In total committees were composed of 2,467 evaluators. Female members tend to be significantly younger and significantly less experienced than male members. There are no significant differences between the relative rank of male and female evaluators.

5 Empirical analysis

Our empirical analysis is structured as follows. First we investigate whether the gender composition of committees matters. For this we use committee-level data from 1987 to 2007. In particular we test whether the number of male and female candidates who is hired is affected by the gender composition of the evaluation committee. Second, we explore whether this effect has varied over time, and whether

³⁶In most Corps of the Spanish Administration individuals are assigned a ranking based on their seniority, their performance in entrance exams and, sometimes, their performance in internal exams.

it depends on the degree of feminization of the position. Third, we analyze whether the potential gender bias of committees is capturing the effect of other observable characteristics of evaluators. These characteristics are experience, age and ranking. Fourth, we study the way in which the gender composition of committees affects candidates’ outcomes. Specifically, we test whether the gender composition matters because different committees evaluate differently, or because their composition affects candidates’ performance. Fifth, we study whose evaluation is affected by bias? Is it female candidates’, is it male candidates’, or is it all candidates’? Finally, we analyze which committees discriminate. For this we take advantage of the information provided by the preliminary multiple choice test that has been held since 2003 at the beginning of the judge and prosecutor exam, and since 2006 for the court secretary exam: in as much as the mark in this test can be considered a gender-unbiased proxy of candidates’ true quality in the oral tests, it will provide information regarding which committees are gender biased.

5.1 Does the gender composition of committees matter?

In order to test whether the gender composition of committees has an effect on female and male candidates’ chances of success, we exploit the variability in the results obtained by candidates who entered the same exam but, because of the random assignment, were evaluated by different committees. Therefore we focus on the following specification at the committee level:

$$y_{ce} = \alpha_e + \beta gender_{ce} + \delta positions_{ce} + \varepsilon_{ce} \quad (1)$$

where e denotes an exam (i.e., “public exam for judge positions held in 1995”) and ce denotes a certain committee and exam (i.e., “committee number one in the public exam for judge positions held in 1995”). Following this notation, y_{ce} is a measure of how successful were candidates who took exam e and were evaluated by committee c ; α_e is an exam fixed effect; $gender_{ce}$ denotes a measure of the gender composition of committee c in exam e and, finally, $positions_{ce}$ is the (log) number of positions that were initially assigned to committee c in exam e which, because of indivisibilities, could vary across committees.

In particular we consider four different dependent variables, in all cases measured at the level of committee and exam: (1) the (log) number of female candidates who were hired, (2) the (log) number of male candidates who were hired, (3) the share of female candidates among successful candidates and, finally, (4) the (log) number of candidates who were hired, irrespective of their gender.³⁷ Our first two variables measure, respectively, the relative probability of success of female and male candidates who have been assigned to a certain committee. Note that analyzing both the effect on female and on male candidates is necessary since the number of positions that each committee can assign is not fixed: committees may always leave some positions vacant and thus public examinations are not a zero-sum game in terms of the number of male and female candidates that can be hired. Our third measure, the share of female successful candidates among all successful candidates, combines the

³⁷A logarithmic specification has been chosen for the dependent variable in order to account for relative changes in the size of exams. Results are qualitatively unchanged if a linear specification is used instead.

information of the two previous measures. Given the well known problem of estimating regressions with a proportion as dependent variable (Papke and Wooldridge 1996), in our regressions we use the standard transformation of the proportion, $\log \left[\frac{\text{proportion}}{1-\text{proportion}} \right]$. Finally, our fourth variable, the log number of successful candidates, captures whether some committees are relatively more benevolent.

An important methodological point must be noted. The inclusion of exam fixed effects, α_e , makes it possible to capture any factor that, at the exam level, may affect our dependent variables, such as the relative quantity and quality of male and female candidates taking that exam. The existence of a random assignment ensures that the quality and quantity of candidates of each gender that have been allocated to a certain committee is not related to the gender, or any other characteristic, of evaluators [$E(\text{gender} \cdot \varepsilon / \alpha) = 0$]. We shall look further into the randomization of the allocation of candidates to committees in Section 5.5.1.

As explained above, the number of candidates obtaining a position in a given committee may have an effect on the number of candidates obtaining a position in another committee within the same exam (e.g., vacancies can be transferred across committees within a given exam). For this reason, throughout our committee-level analysis we cluster standard errors at the exam level.³⁸

In Table 7 we present results from running regression (1). A greater number of women in a committee is associated with a significantly lower number of successful female candidates (column (1)) and a (correspondingly) significantly greater number of successful male candidates (column (3)). In quantitative terms, given that each committee is on average composed of eight evaluators, our results suggest that each additional female member in the committee decreases by 2.8% the chances of female candidates and increases by 3.9% the chances of success of male candidates. In terms of standard deviations, an increase in one standard deviation in the number of women in the committee is associated with a decrease of 3.9% in the number of female successful candidates, and an increase of 5.4% in the number of male successful candidates. Consistently, the number of female evaluators in the committee also has a significantly negative effect on the percentage of successful candidates who are female (column (5)). In column (7) we check whether the gender composition of the committee has any significant effect on the total number of candidates who were hired; the data lead us to reject this hypothesis. Finally, the size of the committee in terms of the number of positions that have been allocated to it is, as expected, positively correlated both with the number of female successful candidates, the number of male successful candidates, and the total number of successful candidates; though it has no effect on the proportion of successful candidates who are female.³⁹

5.1.1 Nonlinearities

There are no clear theoretical reasons to expect that the effect of gender composition is linear. Therefore we next consider the possibility that there are nonlinear effects. We classify committees according to their share of women in six groups as described above: (i) less than 10% of committee members

³⁸Running the regressions without clustering standard errors yields similar results. Results are available upon request.

³⁹About 10% of the evaluators initially rostered are replaced. Given that the random allocation between committees and candidates is with respect to the original composition, we have also instrumented for the final composition of committees using the original composition of committees. Results are available upon request.

are women (or, equivalently, committees with no woman); (ii) there are at least 10% of women, but less than 20% (committees with one woman); (iii) there are at least 20% of women but less than 30% (committees with two women); (iv) there are at least 30%, but less than 40% of women (committees with three women); (v) there are at least 40%, but females are not a majority in the committee; (vi) there is a female majority, that is, over half of members are female. In columns (2), (4), (6) and (8) in Table 7 we present results from running regression (1) using the above four dependent variables, where the omitted gender group is committees with no women. Three things are worthy of mention. First, having at least one female in the committee does matter. Male candidates are 16% (significantly) more likely to succeed in committees with one female evaluator compared to committees where all evaluators are male. The effect is smaller and of opposite sign for female candidates: they are 8% more likely to succeed when evaluated by an all-male committee, although in this case the effect is not statistically significant. Second, the estimated coefficients for groups (ii), (iii), (iv) and (v) are very similar. That is, once there is a woman in the committee, additional female evaluators do not affect candidates' chances much, as long as female evaluators do not become a majority in the committee. Third, being evaluated by a female majority committee does have a large and significant effect: male candidates have 33% higher chances than if evaluated by no women; female candidates have 18% lower chances than if evaluated by no women.

Figure 5 shows the results obtained in column (6), this is, the relationship between the share of successful candidates who are female and the gender composition of the committee that evaluated them. As this figure illustrates, the bottomline from the above set of results is that there seems to exist three different types of committees in terms of their effect on candidates' probability of success by gender: male-only committees, committees where female members are a minority, and committees with a majority of female members.

In the upper panel of Table 8 we run regression (1) again, now only including these three groups to capture the gender composition of committees. In all cases, the adjusted R-square is higher than that of the specification used in Table 7 (columns (2), (4), (6) and (8)), suggesting that this is a better specification. In light of these results, in what follows gender composition will be captured by these three dummy variables. We perform F tests to check whether the minority and the majority coefficients are statistically different: we can reject that a female minority and a female majority have the same effect on (a) the number of successful male candidates and (b) the share of female successful candidates at the 2 percent level (Table 8, upper panel).

What are the reasons behind such nonlinear effects? The first nonlinearity, the fact that having at least one woman in the committee has a relatively large effect, but that additional female members do not, is at least intuitively consistent with male evaluators changing their behavior when in mixed-gender committees. This could work in two different ways. On the one hand, it is consistent with male evaluators discriminating against female candidates and them being especially likely to do so when sitting on mixed-gender evaluation committees—maybe because mixed-gender committees make gender issues more salient. On the other hand, the evidence is also consistent with men favoring some female candidates when there are no women in the committee, but refraining from doing so whenever there is at least one woman in the committee.

The second nonlinearity, the large effect of having a male or a female majority in the committee, is probably related to the fact that voting is decided on a majority basis. Again, since we only observe the aggregate decision of the committee, the evidence is potentially consistent with two possible explanations. First, it is possible that a female (or a male) majority is able to discriminate against candidates according to their gender in a way that just a few women (or men) in a committee cannot. Alternatively, it could be that evaluators react more strongly when their gender is a minority within the committee.

5.1.2 How did it evolve over time?

Next we explore whether the observed effect of the gender composition of committees has changed over time. In order to investigate this we split our sample into two subsamples by decade, 1987 to 1996 and 1997 to 2007, and then run regression (1) on each subsample. As shown in the mid and lower panels in Table 8, the effect of gender composition has not significantly changed over time; if anything, it has increased slightly. In both subperiods, a female (male) candidate is less (more) likely to succeed if she is evaluated by a committee with a female majority than if she is evaluated by a committee with no female. A similar result is found for committees where at least one (but not the majority) of candidates is female compared with committees with no females, although in the case of female candidates who took the exam between 1987 and 1996 this effect is not statistically significant from zero.

5.1.3 Does it depend on the degree of feminization of the position?

Is the effect we find larger whenever the field is very feminized, or is it larger in fields where the incidence of women is low?

Some social psychologists have argued that in order to maintain a positive social identity, women in male dominated fields may tend to identify with male rather than female colleagues (*self-enhancement drive*, Ely 1995). Accordingly, in these fields female evaluators would favor male candidates (Graves and Powell 1995, Goldberg 2005).

We now exploit the fact that the Corps we study here show substantial variability in their degree of feminization. The notary Corps, with an average of 13.3% of female members during the period considered, is relatively masculine. The court secretary Corps, where almost 53.4% of members were female, is relatively feminine. In between, the judge and prosecutor Corps are intermediate cases, with 30.4% and 36.5% of females amongst members respectively.⁴⁰

If the *self-enhancement drive* hypothesis is correct, one would expect the effect we find to be greater in exams for positions that are relatively masculine. Table 9 tests this hypothesis by running regression (1) by type of examination. Focusing on column (3), displaying the effect on the proportion of successful candidates who are female, the estimated coefficients are similar for most examinations.

⁴⁰The judge and prosecutor exam represents years 2001-2007 and thus shows a larger incidence of women.

Only in the case of the prosecutor examination, an intermediate case in terms of feminization, do we find a smaller effect. Thus we do not find evidence supporting the *self-enhancement drive*. Indeed, it does not seem that the effect depends on the degree of feminization of the field.

5.2 Is it due to other committee characteristics?

The message from the above tables is clear: female (male) candidates are significantly more likely to succeed if they are randomly assigned to committees in which there are relatively more male (female) evaluators. However, differences in the gender composition of committees can be associated with differences in other committee characteristics. Indeed, Table 6 shows that, consistent with the later entry of women in the Judiciary, male committee members tend to be older and more experienced than female members. Therefore, being assigned to a committee with more women also implies being assigned to a committee that is younger and less experienced. This leads to obvious concerns: are the observed findings driven by the gender composition of committees, or are they due to some other committee characteristic?⁴¹

To tackle this issue we introduce as controls in regression (1) the following committee characteristics: the mean age of committee members, their mean ranking, and their mean experience as committee evaluators. Results are displayed in Table 10.

Male candidates tend to be relatively less successful when they are assigned to committees whose members are more highly ranked (column (2)). In quantitative terms, an increase in one standard deviation in the ranking of committee members reduces the chances of male candidates by 6.2%.⁴² Female candidates fare slightly better when evaluated by more highly ranked committees but this effect is not significantly different from zero (column (1)). Neither age nor experience of committee members seems to play a role in candidates' success by gender.⁴³

The effect of the female composition of the committee is not significantly affected by the inclusion of these additional controls; thus it does not seem that our previous results were caused by these committee characteristics.

5.3 How does it matter: differences in performance or differences in evaluation?

In this section we investigate further the nature of the bias that we find. In our setting, we lack a proper *placebo*: candidates taking the exam observe, and as a matter of fact, know beforehand, the exact gender composition of the evaluation committee. This could, in principle, affect candidates' behavior in two different ways. First, once they observe the composition of the evaluation committee

⁴¹For example, there seems to be some evidence that younger referees are harsher (Cole *et al.* 1978, as cited by Broder 1993).

⁴²The standard deviation in the average relative ranking of committee members equals 0.12 (Table 4).

⁴³The results from this table do not change if we use a linear specification, or if we use the nonlinear specification in Table 7 with groups (i) to (v).

some candidates may decide not to take the exam. However, the nature of these examinations suggests that this is not likely to be the case. Indeed, in order to have some chance of success candidates need to prepare full time for several years and it is thus very unlikely that someone would ultimately decide not to take the exam because of the gender of her evaluators.⁴⁴ Second, candidates' performance in the exam could be affected by the committee gender composition. This would be in line with Claude Steele's "stereotype threat" hypothesis, according to which female candidates may perform more poorly on an exam when they are reminded of their gender (Steele 1997).

Therefore we cannot rule out ex-ante the possibility that the gender differences in success rates we observe reflect gender differences in performance on the exam induced by the gender composition of the committee; and thus have nothing to do with discrimination or biases by committee members.

In order to learn more about this hypothesis, we now compare the bias in exams first written by the candidate and then read to the committee, with the bias in exams where the entire process was oral. As described in the background section, exams for judge, prosecutor and court secretary positions used to be written and then read by the candidate in front of the committee until approximately 1995, when a legal reform introduced completely oral testing. It seems reasonable to think that in the former case there should be less room for candidates' performance being influenced by committee composition. If the effect we observe is performance based, then we would expect it to be smaller for written exams. It should be noted that the structure of the exam is such that the possibilities for candidates' performance being affected by the gender (or characteristics) of evaluators are limited also in oral tests. Tests are based on learning and reciting legislation and the choice of topics is decided by a random draw. Moreover, as mentioned before evaluators and candidates seldom interact during the test.

As shown in Table 11, the effect of gender composition is very similar in both subsamples and there are no significant differences in the estimated coefficients, suggesting that the gender composition of committees matters through evaluation and not through performance.

5.4 Who is discriminated against?

The analysis above shows the existence of a gender bias in evaluation. Indeed, we have found that female candidates are more likely to be hired when there are fewer female members in the evaluation committee. At the same time, male candidates are more likely to be hired when there are more female members in the committee. However, note that it would be incorrect to interpret this result as providing evidence of both male and female candidates' evaluations being affected by the gender composition of committees. As pointed out above, when the number of candidates that passed every stage is larger than the number of positions available, positions are assigned based on candidates' ranking. Therefore, it is at least possible to think about a setup where the gender composition of committees only affects the evaluations received by candidates of a certain gender but where,

⁴⁴Evidence from multiple choice exams, where attendance is reported, suggests that in fact a few candidates do not show up, but this is usually due to the fact that the official registration has to be done several months in advance, usually when the previous year's public exam has not even finished.

due to the limited number of positions available, the chances of success of candidates of the other gender are also affected indirectly, as their position in the ranking is affected. That is, given that the number of positions that can be assigned in an exam has an upperbound, our evidence is in principle consistent with three different scenarios: (1) the gender composition of committees affects the evaluations obtained by female candidates, (2) the gender composition of committees affects the evaluations obtained by male candidates, and (3) both female and male candidates' evaluations are affected.

We analyze who is discriminated against using two approaches.

First, the existence of exams where not all positions have been assigned allows us to gain more insight into this issue. In exams where all positions have been assigned, whether a (male or female) candidate obtains a position depends on the evaluations obtained by other candidates; this is not the case in exams where the number of vacancies is larger than the number of candidates who have passed every test. Therefore while in exams where all positions have been assigned, a comparison across committees does not provide us with information on who is being favored or discriminated against; in cases where all initial positions were not assigned (i.e. where some vacancies were left), a comparison across committees will tell us whether some positions are not being assigned to women or men,⁴⁵ independently of performance by the other gender.

Therefore we now focus on the exams in which some positions were not assigned; in these cases, was it female or male candidates who were left without a position? As it turns out, in these exams female candidates are not significantly affected by the gender composition of the committee (see column (1), Table 12). However, the probability of success of male candidates is significantly greater when there are more women in the committee (column (2)). Comparing these results with those obtained with the whole sample (in Table 10), the estimated coefficient for female candidates decreases by half when we only consider cases where not all positions were assigned; the estimated coefficient for male candidates is almost unchanged. The evidence is thus consistent with the evaluation of male candidates being directly affected by the gender composition of committees. On the other hand, the evaluation received by female candidates does not seem to be directly affected.

Second, we make use of the availability of exam grades for successful candidates. Information from successful candidates' exam grades needs to be taken with caution: these grades only reflect the evaluations received by the upper part of the distribution (around 5% of all candidates); nevertheless they provide us with complementary information to the analysis above. Note also that if we compare the grades of all successful candidates across committees a clear sample bias problem would arise, as the number of people who succeed in a committee is determined by the relative grading standards of the committee (e.g., a lower average grade could be due to more candidates, albeit of lower quality, having passed). In order to avoid this problem, we construct two samples, one with female candidates and one with male candidates, where we only include the first N_e^f female and the first N_e^m male candidates in each committee respectively. We denote by N_e^f and N_e^m the minimum number of female and the minimum number of male candidates who obtained a position in some committee in exam

⁴⁵Note that as empty positions in one committee are allocated to any other committee that needs them, so long as there is at least one vacant position left at the exam level, any committee can potentially assign one more position.

e. For instance, if a given exam was composed of three committees where, say, 25, 26 and 27 female candidates respectively obtained a position, we exclude from our female sample female candidates who ranked 26th in the second committee; similarly we exclude the two female candidates who ranked 26th and 27th in the third committee. Restricting our sample to these two sets that include, respectively, the best N_e^f and N_e^m female and male candidates of each committee, reduces our female sample from 4,561 to 3,337 observations, and decreases the size of the male sample from 3,139 to 2,145 observations.

In order to find out whose grades are affected by the gender composition of committees, in Table 13 we examine how the characteristics of the evaluation committee affect the grades obtained by successful candidates, where grades have been normalized between zero and one. As in previous results, we focus on a specification where gender composition is summarized by nonlinear variables: whether there is a minority, and majority, of female evaluators in a committee. We also control for committees' average age, experience and ranking. There exists the possibility that one candidate's grade is affected by the performance of another candidate evaluated by the same committee; for this reason we cluster standard errors at the committee level.

Being evaluated by a committee with a minority of female evaluators implies a (marginally) significant increase in the final grade obtained by males of two points, out of 100 (column (1)), but has no significant effect on females' grades (column (2)). Likewise, being evaluated by a female majority committee implies a (marginally) significant increase in males' final grade of five points; there is no significant effect on females' grades. These results are thus consistent with those obtained in Table 12 by looking at the number of successful candidates, in that it is only male candidates' evaluation that is being affected by the gender composition of the committee.

5.5 Multiple choice test evidence

We have found that the gender composition of committees affects candidates' chances of success: in particular, we have found that male candidates tend to receive a relatively higher (lower) evaluation if they are assigned to committees composed of more female (male) evaluators. However, from that evidence we cannot tell whether this is due to male dominated committees being excessively tough with male candidates, or to female dominated committees being too benevolent with them. In order to test which committees are biased we need to control for candidates' true quality. For this we are going to use multiple choice test information as proxy for quality. Additionally, the multiple choice test information might be useful in order to check that the assignment of candidates to committees was indeed random.

Our multiple choice test information comes from two examinations. In addition to the two oral tests, a preliminary qualifying multiple choice test has been introduced in the exam for judge and prosecutor positions (since 2003), as well as in the exam for court secretary positions (since 2006).

In these exams in total the evaluation was performed by 30 committees which evaluated a total of 10,329 candidates.⁴⁶ Each committee was composed of nine members and the number of female

⁴⁶This figure excludes the 14,201 candidates who failed the multiple choice test and thus were not assigned to evaluation

evaluators ranged between a minimum of one and a maximum of five.

5.5.1 Is the allocation of candidates to committees random indeed?

As explained in Section 3.2.1, candidates to positions in the Judiciary are assigned to committees through a lottery which determines the candidate initial to start the matching. After the lottery, the list of candidates that have been assigned to each committee is made public, which allows the public to verify that the assignment has been random. Nonetheless, here we want to formally test that this allocation was random by taking advantage of the multiple choice data available. We do so in two ways and provide results in Table 14.

First, we check whether the gender of the candidate is related to the gender composition of the committee who evaluated the candidate: we regress a dummy variable for female candidates on our set of gender decile composition dummies (ii), (iii), (iv), (v) and (vi) used above for committee membership by gender (given that there are nine committee members in the exams for which a multiple choice test exists, these dummies are equivalent to dummies for committees with one, two, three, four and five female members, respectively). Second, if the assignment is random, we would not expect to see statistically significant differences in the quality of female candidates that have been assigned to different committees. The same applies for male candidates.

Results suggest that the assignment is indeed random. None of the 12 coefficients estimated is significantly different from zero at the five percent level. One coefficient is significantly different from zero at the ten percent level: the quality of female candidates assigned to committees with four women is marginally lower than the quality of female candidates who were assigned to committees with only one women.

5.5.2 Which committees discriminate?

The evidence presented in the previous section suggests that the gender composition of committees affects candidates' chances of success. In particular, we have found that male candidates tend to receive a relatively higher (lower) evaluation if they are assigned to committees composed of more female (male) evaluators. Nonetheless, the above evidence does not allow us to disentangle whether this is due to male dominated committees being excessively tough with male candidates, or to female dominated committees being too benevolent with them. In order to test which committees are biased we need to control for candidates' true quality. We turn to this issue in this section by making use of the multiple choice test information explained above.

The material required for the multiple choice test is contained in the material that is due for the oral committees.

In 2006, a new rule was introduced according to which candidates who had passed the multiple choice test but failed the oral stages the previous year were exempted from retaking it again. For this reason, the figure also excludes 980 candidates for judge and prosecutor positions who in 2006 and 2007 were exempt from the preliminary multiple choice test.

stages of the examination; the mark obtained in the multiple choice test can thus be a good measure of candidate quality in the oral stages.⁴⁷

In assessing the connection between multiple choice testing and gender, we can turn to a large literature that investigates whether men have a comparative advantage in multiple choice tests. First, the empirical evidence does not in any case point towards women having a relative advantage in multiple choice tests. Rather, several recent studies have found no significant gender differences among economics tests using fixed response tests rather than constructed response tests (Walstad and Becker 1994, Greene 1997, Chan and Kennedy 2002). If that is the case, then the information from our multiple choice test can be considered gender unbiased. Finally, some studies (Bell and Hay 1987, Lumsden and Scott 1987, Bolger and Kellaghan 1990) have found that males may have a relative advantage on multiple choice tests. In that case, we need to take this into account in order to interpret any potential bias.

First we test whether the results found in the previous section also hold in a sample that includes all candidates, both successful and unsuccessful. Here we are able to use candidate-level information. In Table 15 we display the results from running the following tobit regression:

$$y_{ie} = \alpha_e + \sigma \text{candidate}_i^f + \varphi \text{evaluators}_{ce}^f + \eta \text{candidate}_i^f * \text{evaluators}_{ce}^f + \phi \text{test}_{ie} + \varepsilon_{ie},$$

where y_{ie} denotes the final exam grade of candidate i obtained in exam e , candidate_i^f is a dummy variable equal to one in case that candidate i is a woman and zero otherwise, and test_{ie} denotes the mark obtained by candidate i in the multiple choice test in exam e . We use a tobit specification with left-censoring at the 25 marks threshold (as we only observe grades for candidates who pass both oral exams, each one requiring a minimum grade of 12.50). As our measure of gender composition of the committee we follow the previous analysis and allow for nonlinear effects, captured by evaluators_{ce}^f , denoting the set of gender decile composition dummies for committee c and exam e . Since every committee in our sample has at least one female member, now the control group is committees with one woman only. Given that there are nine committee members, the set of gender decile composition dummies (ii), (iii), (iv), (v) and (vi) used above is in this case equivalent to committees with one, two, three, four and five female members, respectively.

Again, given the random procedure followed to allocate candidates to committees within each exam, significant differences in the characteristics of candidates who have been allocated to different committees are not to be expected [$E(\text{evaluators} \cdot \varepsilon | \alpha) = 0$]. In fact, the process of random allocation of candidates to committees actually yields groups of candidates that are not significantly different in either their gender composition or their quality as measured by their test performance. As mentioned before, the multiple choice test mark can thus be expected to be either independent of gender, or biased upwards for male candidates relative to female candidates [$E(\text{candidate} \cdot \varepsilon | \text{test}, \alpha) \geq 0$].

⁴⁷The multiple choice test consists of 100 questions in both exams. Each question lists a set of four possible answers. If the answer to a question is correct, the candidate receives one mark. If the answer to the question is incorrect, the candidate loses 0.33 marks (and therefore the expected value of a randomly answered question is zero). If the question is left unanswered, the candidate gets zero marks. For details on the material due for each test and duration see the appendix.

Results are shown in column (1) in Table 15. The multiple choice mark is positively and very significantly associated with the final grade. The results also reveal the source of the gender bias previously found: while in committees where female evaluators are minority the grades obtained by male and female candidates do not differ, in committees with a female majority male candidates obtain an estimated 3.86 more marks than female candidates who had obtained a similar evaluation in the previous multiple choice test. Results are qualitatively similar if, instead of the grade, we use as the dependent variable whether the candidate obtained the position or not (as seen with the probit specification in column (2)).^{48,49}

Putting together these two pieces of evidence, our results suggest that female majority committees are overestimating the true quality of male candidates. In the case that the format of the test is gender neutral, we can also say that female minority committees are not biased. Alternatively, in the case that men have a comparative advantage in multiple-choice tests, the bias of female committees in favor of male candidates we find above would be underestimated; at the same time, male committees might be favoring male candidates too.

6 Discussion

In this section we summarize our results, and discuss the potential limitations of our analysis and possible behavioral interpretations.

6.1 Summary of results

In this paper we find that the gender composition of a committee is an important determinant of success in public examinations. In particular, male candidates are relatively more likely to succeed when they are (randomly) assigned to a committee with a larger number of female evaluators; female candidates, on the other hand, are relatively less likely to succeed in such circumstances. The effect of the gender composition of committees is highly nonlinear. First, having at least one female in the committee has a great impact on candidates' chances of success, suggesting that the female member's presence could be affecting the voting behavior of male evaluators. Male candidates are 16% (significantly) more likely to succeed in committees with a female minority compared to committees where all evaluators are male. Analogously, female candidates are 10% more likely to succeed when evaluated by an all-male committee, although in this case the effect is not statistically significant. Second, once there is a woman in the committee, additional female evaluators do not affect candidates' chances much, as long as female evaluators do not become a majority in the committee. Third, being evaluated by a female majority committee does have a large and significant effect: male candidates have 34% higher chances than if evaluated by no women; female candidates have 17% lower chances than if evaluated by no women (Table 8, upper panel).

⁴⁸Controlling for other committee characteristics leaves results unchanged.

⁴⁹Given the small number of groups (30 committees), standard clustering does not provide a consistent estimation of standard errors (Wooldridge 2003). In our case clustering actually reduces standard errors; qualitatively results are the same.

These results are not specific to one particular decade (Table 8, lower panels), and do not depend on the degree of feminization of the position (Table 9). Furthermore, our results are not due to omitted characteristics of evaluators such as ranking, experience or age (Table 10). Indeed, while male candidates fare worse when evaluated by more highly ranked committees, including ranking as regressor does not affect the aforementioned gender effect.

In line with Claude Steele’s “stereotype threat”, it could be that some of the effect we are detecting is due to differential performance by candidates as a reaction to the gender composition of the committee. In order to test this hypothesis, we have exploited the fact that in some years exams were first written by the candidate and then read in front of the committee. While in these exams there should be even less interaction between committees and candidates than in oral exams, we still observe a gender bias of similar size, suggesting that our results seem to be due to evaluators’ perceptions rather than different performance by candidates depending on the gender of their evaluators (Table 11).

Looking at the evidence from committees where not all the initially available positions were filled, and exploiting the information provided by the grades obtained by successful candidates, we have then shown that it is male candidates’ evaluations that are being directly affected by the gender composition of the committee (Tables 12 and 13). The effect on female candidates is only indirect and operates via their relative position in the exam ranking: they are only affected by the gender composition of committees in those exams where there were more good candidates than positions. In those cases, the fact that male candidates had received a relatively high grade implied that some female candidates ended up not being high enough in the ranking and, therefore, not getting a position.

The evidence from multiple choice tests further reveals that, even though the quality of female and male candidates does not vary across committees (Table 14), in female dominated committees male candidates obtain about four more marks than female candidates who had obtained a similar grade in the preliminary multiple choice test (Table 15, column (1)); in these committees male candidates also enjoy a significantly higher probability of being hired (Table 15, column (2)). Thus, in essence our results show that female majority committees are favoring male candidates.

6.2 Behavioral interpretation

The empirical analysis shows that the gender composition of committees matters for candidates’ success, and that in particular, committees with a female majority seem to favor male candidates in their evaluation. Unfortunately, we do not observe individual votes within committees, but only the committee outcome. That means that we cannot directly attribute discrimination to female or male committee members. It could be the case that female members in the committee, who might tend to favor male candidates, can only do so whenever they reach majority in a committee (due to majority voting) or it could be that men in committees favor male candidates when sitting on mixed-gender committees (maybe because mixed-gender committees make gender issues more salient).

Moreover, we can only speculate about the possible explanations for the effect we find, as this may be as much in the domain of psychology as of economics. Nevertheless, below we discuss the potential

explanations for the existence of a gender bias and point out which of them seem more consistent with the observed evidence.

6.2.1 Taste discrimination

Models based on taste, such as Gary Becker’s (1957) seminal work, suggest a preference-based motivation for the existence of discrimination. Our results are consistent with positive taste discrimination in favor of male candidates by female majority committees. There are several explanations whereby evaluators may have different objective functions.

First, perhaps committees are more generous towards the opposite gender out of political correctness. But this seems at odds with the fact that it is male candidates who are favored here. Second, perhaps this is a case of beauty in the labor market: in the spirit of Hamermesh and Parker (2005) and Hamermesh and Biddle (1994), female evaluators might feel sympathetic towards candidates of the opposite gender, particularly when they are good looking. Again, according to our results this would be an explanation in case it is male beauty only that counts.⁵⁰ Unfortunately, no proxy for candidate beauty is available in the data set. It should also be noted that there is a wide age gap between committee members and candidates, such that successful candidates are not likely to interact with committee members later in their careers.

Third, one hypothesis in the field of social psychology is the so-called *similarity attraction* paradigm (Byrne 1971). According to this hypothesis, evaluators might favor candidates with similar characteristics or attitudes. However, this is not consistent with the observed evidence: female candidates obtain relatively higher grades when they are evaluated by an all-male committee.

Fourth, social psychologists have also argued that in order to maintain a positive social identity, women in male dominated fields may tend to identify with male rather than female colleagues (*self-enhancement drive*, Ely 1995). Accordingly, in these fields female evaluators would favor male candidates (Graves and Powell 1995, Goldberg 2005). In our framework we observe exams for positions with very different degrees of feminization. As shown in Table 9, the effect of the gender composition of committees on chances of success does not depend on the type of examination; and thus does not depend on the degree of feminization of the position. That is, it does not seem that the *self-enhancement drive* is at work here.

Finally, social norms could potentially be behind our results. The gender of candidates might enter evaluators’ objective functions if there is societal discrimination: this would be the case if there exists a social norm dictating that members of the Judiciary should be male; however it is not clear why only female evaluators would be affected by such social influences.⁵¹

⁵⁰Hamermesh and Parker (2005) find that better looking professors, and especially male ones, are more highly rated by students. Hamermesh and Biddle (1994) find that attractive individuals, especially males, earn higher wages than the unattractive.

⁵¹Bertrand *et al.* (2005) explore the concept of *implicit discrimination*: it could be that, in a more subtle version of taste discrimination, perhaps due to time pressure, female evaluators end up favoring male candidates due to some unconscious mental association between a target (male) and a given attribute (being a good member of the Judiciary), or even due to some unconscious lack of confidence, or complex. Our context, however, does not fit that framework:

6.2.2 Statistical discrimination

Statistical models of discrimination argue that, in the presence of information asymmetries about the real productivity of individuals, their gender can be considered a signal that provides additional information. Genders may differ in their expected productivity (Phelps 1972, Lazear and Rosen 1990), in the reliability of observable signals (Aigner and Cain 1977, Borjas and Goldberg 1978, Cornell and Welch 1996) or in their relative quality across different dimensions (Bagues and Perez-Villadoniga 2005).

The results in this paper are consistent with a situation in which female evaluators overestimate the quality of male candidates (regardless of the fact that this belief might be wrong), while male evaluators do not hold gender-based beliefs. That is, it is possible that female evaluators believe (even if incorrectly) that men are somewhat more skilled than they really are; thus female evaluators take the gender of candidates as a signal of their abilities, which leads to an overestimation of male candidates' quality.

Our results could explain why women shy away from competition (Niederle and Vesterlund 2007), or that women do not perform as well as men when competing against men (Gneezy *et al.* 2003). If women suffer from a *complex* that makes them overestimate men's quality, that could at least partially explain women's reluctance to compete and their poorer performance when competing against men.

6.2.3 Strategic discrimination

Finally, our results could be due to strategic behavior. If there exists uncertainty about the quality of female and male members of the Judicial Corps, but the average quality of members is publicly known, committee members might want to increase the average quality of their group. This would be consistent with women being tougher on female entrants (and men being tougher on male entrants).⁵² Given the dramatic increase in the incidence of women in the Judiciary in recent times, we would expect this effect to have decreased over time; however results in Table 8 suggest the effect has not decreased.

7 Conclusions

In order to remedy the historic underrepresentation of women in decision-making positions, many countries have begun encouraging or mandating gender parity at the top levels of the public and the private spheres. The motivation underlying the imposition of gender parity is the existence of the so-called 'glass ceiling'—beyond which women cannot go any further—and the perception that it might be due to (male) discrimination against women. If women are not currently able to break the

committee members in public examinations get to evaluate candidates during extended periods of time (over an hour), so time pressure is not really an issue.

⁵²This does not take into account that both groups would end up increasing the quality of the whole group; given uncertainty, the situation would be consistent with rational choice.

glass ceiling, imposing gender parity at the top level should increase hiring of other women, and in turn increase the incidence of women in decision-making.

This could work at least in two different ways. On the one hand, there could be indirect effects, such as the influence of role models on young women’s career decisions. On the other hand, women who get to top-level positions because of gender quotas might hire more women than their male counterparts. This will be the case if female candidates are more likely to be recruited when evaluated by female evaluators. Our data, however, suggest that the latter is not necessarily true.

This paper uses data from 51 Spanish public exams to the main Corps of the Judiciary between 1987 and 2007 involving approximately 150,000 candidates and 2,467 evaluators. The exams provide a rich source of information because of the characteristics of the evaluation process: first, candidates are allocated to committees randomly, which automatically eliminates concerns about endogeneity; second, the subjects and the experiment are actually taken from real life, hence avoiding the usual problems associated with artificial settings; finally the experiment is relevant because of the importance, and magnitude, of public examinations in Spain and other countries.

The main result in this paper is that male (female) candidates allocated to female majority committees are relatively more (less) likely to be hired. This effect is robust to the particular specification used, the subperiod considered, the degree of feminization of the Corps, and the inclusion of committee characteristics as controls. Furthermore, we show evidence against the interpretation of this effect as due to a ‘stereotype threat’. Moreover, the availability of an objective measure of quality allows us to disentangle the source of the bias: we find that female majority committees overestimate the quality of male candidates.

What is the explanation behind the bias we find? We can only speculate about the answer to this question, but our results are more consistent with two hypotheses: first, a form of statistical discrimination whereby female evaluators suffer from some complex that leads them to believe that male candidates are of higher quality than their real quality. If women suffer from some inferiority complex that makes them overestimate men’s quality, that could at least partially explain women’s poorer performance when competing against men, as well as their lower propensity to compete, as found in Gneezy *et al.* (2003) and Niederle and Vesterlund (2007) respectively. Second, our results are also consistent with male members in a committee changing their vote as a function of whether there are women in the committee. In particular, men in committees might be favoring male candidates when there are women in committees, perhaps because in that case male committee members’ identity is strengthened (Akerlof and Kranton 2000).

Our findings have direct policy implications for Spain, where the government has just passed the so-called Equality Law to tackle discrimination against women.⁵³ Such legislation imposes gender parity in all recruiting committees, including the committees we analyze here. However, this paper suggests that such a policy will be not only ineffective but even counterproductive: a simple back of the envelope calculation shows that, had there been an additional woman in every committee for the exams we study here, 123 fewer women would have succeeded, representing—approximately 2.8% of

⁵³Equality Law, BOE, 71, March 23, 2007, p. 12611.

females hired.^{54,55}

Instead, a deeper understanding of the factors limiting women’s participation should be gathered in order to derive adequate policy. In the Spanish case, a quick glance at *résumés* of the members of the first parity cabinet shows the difficulty in balancing family and career: while the eight male ministers have 24 children in total, the eight female ministers only have five in total (*El País*, October 16, 2005); which is consistent with Lazear and Rosen (1990).⁵⁶

Last, but not least, our evidence also has implications for the general design of public examinations. At the moment, most public exams in Spain are in oral form: even exams in written form have to be read by candidates in front of committees; therefore anonymity is lost. According to the evidence in this paper, introducing anonymity in public examinations would improve the fairness of the evaluation.

8 Appendix

In the following appendix we describe, for every type of examination, the rules regarding committee composition and format of exams.

8.1 Judge examination

For the judge exam, in place until 2000, committees were presided over by the President of the Superior Court of Justice or a Judge of the Superior Court of Justice. For the rest of the committee members, we have to distinguish between the periods before 1995, between 1995 and 1998, and between 1998 and 2000. Before 1995, in addition to the president, evaluation committees had to be composed of one judge, two professors in law, one prosecutor, a public defender and a lawyer (*Ley Orgánica 6/1985, July 1*). Since 1995 and until 1998 the rest of committee members were to be one judge, one prosecutor, two law professors in two different fields, one professional lawyer, and another judge, who was to be the committee’s secretary. From 1998 until 2000, the rest of committee members had to be

⁵⁴The calculation has been done as follows. Given that on average each committee is composed of eight members, the results reported in the first column of Table 7 indicate that including an additional female evaluator in each committee would have decreased by 2.8% the number of female successful candidates. That is, approximately 123 fewer women would have obtained a position (out of a total of 4,561 who did succeed).

⁵⁵The law has other negative unintended effects. Since committee members must have a certain minimum qualification, the few females who are eligible to sit in evaluation committees actually end up spending proportionally much more time in committees than their male colleagues. The drawbacks of this policy are obvious in fields such as academia, where sitting in committees takes time away from research.

⁵⁶Our results are also consistent with anecdotal evidence from the Spanish cabinet. The introduction of equal numbers of male and female ministers in 2004 (from only three women out of 16 ministers in previous years) has not been followed by significantly higher numbers of women in subordinate top public positions: 18.9% of general directors were women in 2006, compared to 15.8% in 2002 with the then-ruling Popular Party; an increase that is consistent with the trend. Similarly, 24.1% of deputy secretaries and undersecretaries were female in 2006, while this figure was 16.8% in 2002. That is, newly appointed female ministers have not hired significantly more women than their (mostly male) predecessors.

Another event from Spanish recent history is also consistent with our findings. With the advent of the Second Republic in 1931, despite not being allowed to vote, women were allowed to run for elections for the first time. In these first democratic elections, three women were elected: Clara Campoamor, Victoria Kent and Margarita Nelken. That same year, when the Parliament discussed whether women should be allowed to vote, out of the three female MPs only Campoamor turned out to support feminine suffrage. However, a majority of male MPs voted in favor and thus female suffrage was approved.

two judges, one prosecutor, two law professors in two different fields, one lawyer with over ten years of experience, one public defender, a court secretary of first category, and another judge (a member of the technical agencies in the General Council of the Judicial Corps with a law degree), who was to be the committee's secretary. Whenever it was not possible to roster full professors, associate professors were to be considered (*Ley Orgánica 5/1997, December 4*).

The format of the exam has also varied over time. Until 1995 there were two qualifying tests in *written* format, whereby questions had to be first written by the candidate and then read to the committee. In the first test, of theoretical content, the candidate had to answer three questions in six hours; the questions were randomly chosen. In the second test, which was practical, the candidate had five hours to answer two questions which were randomly chosen out of a total of five possible questions proposed by the committee. In both exams a clerk had to be present with a copy of the candidate's answers so as to check that the candidate would not change the content while reading (*Orden January 28, 1987, BOE, 31, February 5, 1987, pp. 3525-31*). In 1995, a reform implemented oral testing (*Acuerdo June 7, 1995, BOE, 166, July 13*). Since then the exam was composed of two qualifying oral tests of theoretical content. In both tests candidates had to answer five randomly chosen questions within 75 minutes.

8.2 Prosecutor examination

For the prosecutor exam, in place until 2000, evaluation committees had to be composed of one prosecutor from the Superior Court of Justice (who was to preside over the committee), two prosecutors, one judge, one professor in law, one lawyer (to be chosen by the Lawyers' Corps), a public defender (affiliated with the Ministry of Justice), and one prosecutor working for the Technical Secretariat in the State's General Prosecuting Office (who was to be the committee's secretary) (*Ley 50/1981, December 30*).

Regulations regarding the format of exams for prosecutor were always similar to those for exams for judge, as specified by Article 42 in *Ley 50/1981, December 30*.

8.3 Judge and prosecutor joint examination

The exams for judge and prosecutor merged in 2001 and since then there has been a joint exam for both positions (*Ley Orgánica 9/2000, December 22*). Committees are composed of nine members. Each committee is formed by a committee president, who must be a top member of the Judicial Corps—either a judge, in the case of oddly numbered committees, or a public prosecutor, in the case of evenly numbered committees; all from a Superior Court of Justice. The other eight members are determined by their respective Corps of origin according to the rules. In particular, the other eight members in the committee must include two judges, two public prosecutors, one law professor, one public defender, and one lawyer with over ten years of professional experience. Finally, the committee secretary is a court secretary of first category.

The exam initially followed the same format used previously both in the exams for judge positions and in the exams for prosecutor positions, that is, it was composed of two qualifying oral tests. Since 2003 a preliminary multiple choice test has been also added (see *Acuerdo* March 20, 2003, BOE, 76, p. 12368).

In 2003 and 2004, the material required for the multiple choice test was the same as in stage two of the examination. Instead, in 2005, 2006 and 2007 the material for the test was slightly changed, as 20 of the questions were on material that is required for the third stage.

The judge and prosecutor multiple choice test had to be answered within two hours during 2003 and 2004, and two and a half hours in 2005, 2006 and 2007. For the sake of illustration, in the 2005 judge and prosecutor exam the distribution of topics covered by the multiple choice test was as follows: ten questions about general theory of law and constitutional law; 40 questions about civil law; 30 questions about penal law; and 20 questions about procedures of law (13 on procedures of civil law, seven on procedures of penal law).

8.4 Court secretary examination

Court secretary exam committees were composed of seven members until 2003 and have included nine members since 2006. Until 1995 committees were presided over by a judge and, as well, were composed of another judge, two court secretaries, one law professor, one lawyer with over five years of professional experience, and a public lawyer (*Orden* January 30 1998, BOE, 45, February 22). Since 1995 the composition is as follows: the president must be a judge, and there must be a second category prosecutor, three court secretaries, one law professor, and one lawyer (*Orden* May 19, 1995, BOE, 140, June 13). Since 2006 there are also one prosecutor and one civil servant of type A.

As in the case of examinations for judge and prosecutor positions, until 1997 exams consisted of two tests in *written* format, whereby questions had to be written and then read to committees; a clerk had to be present with a copy of the candidate's answers so as to check that the candidate would not change the answers when reading. In 1997, a reform implemented oral testing (*Orden* January 21, 1998, BOE, 31, February 5).

Since 2006, the court secretary exam includes a preliminary multiple choice test with the same format and structure as the one in the judge and prosecutor exam. The material for the multiple choice test includes all topics in the first part of the program. The court secretary multiple choice test in 2006 had to be answered within two hours.

8.5 Notary examination

Evaluation committees for notary exams are composed of seven members. In 1987, the president was to be the general director (or one of the subdirectors) of registrars and notaries, or the dean (or vicedean) of the notary college where the examination was to be held. The other six members were

to be: two notaries, a law professor (in the field of either civil, commercial, Roman, international, private, administrative law, or procedures of law), a judge, a registrar, and a state lawyer or a private sector lawyer in the field of civil or commercial law with over 15 years of professional experience (*Real Decreto* 950/1987, July 24). Since 1988, the six members have included two registrars and one notary as opposed to one registrar and two notaries (*Real Decreto* 1752/1987, December 30, *Real Decreto* 1728/1991, November 29).

The notary exam consists of four stages: the first and second stages are oral tests, the third and fourth stages are in written format but again involve candidates reading their written answers in front of the evaluation committee; a clerk has to be present with a copy of the candidate's answers to make sure that the candidate reads literally everything she has written.

References

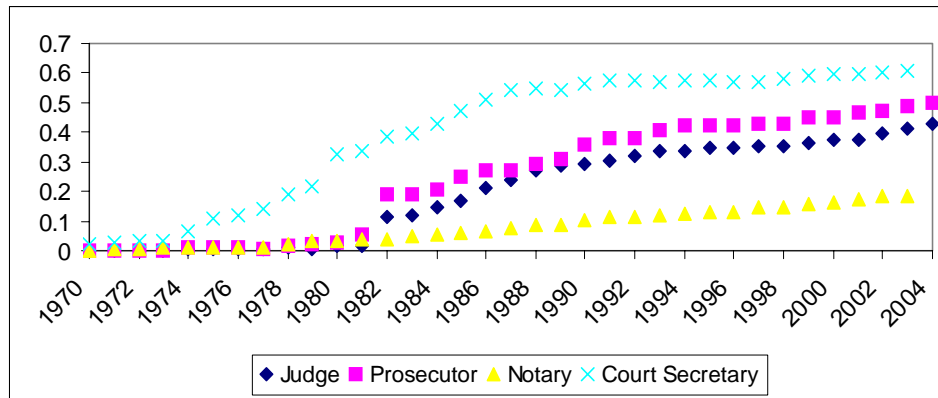
- [1] Aigner, Dennis J. and Glen G. Cain (1977), Statistical Theories of Discrimination in Labor Markets, *Industrial and Labor Relations Review*, Vol. 30(2), pp. 175-87.
- [2] Akerlof, George, and Rachel E. Kranton (2000), Economics and Identity, *Quarterly Journal of Economics*, Vol. 115(3), pp. 715-33.
- [3] Antonovics, Kate, Peter Arcidiacono, and Randall Walsh (2005), Games and Discrimination: Lessons From The Weakest Link, *Journal of Human Resources*, Vol. 40(4), pp. 918-47.
- [4] Ayres, Ian and Peter Siegelman (1995), Race and Gender Discrimination in Bargaining for a New Car, *American Economic Review*, Vol. 85(3), pp. 304-21.
- [5] Bagues, Manuel (2005), ¿Qué determina el éxito en unas oposiciones?, FEDEA Documento de Trabajo 2005-01
- [6] Bagues, Manuel and Berta Esteve-Volart (2006), Gender and Overconfidence in Selection Processes: Evidence from Public Examinations, mimeo, Universidad Carlos III and York University
- [7] Bagues, Manuel and Maria Jose Perez Villadoniga (2008), Why Do I Like People like Me?, mimeo, Universidad Carlos III Working Paper 080601
- [8] Becker, Gary S. (1957), *The Economics of Discrimination* (Chicago: The University of Chicago Press)
- [9] Bell, R. C. and J. A. Hay (1987), Differences and Biases in English Language Examination Formats, *British Journal of Educational Psychology*, Vol 57, pp. 212-20.
- [10] Bertrand, Marianne and Kevin F. Hallock (2001), The Gender Gap in Top Corporate Jobs, *Industrial and Labor Relations Review*, Vol. 55(1), pp. 3-21
- [11] Bertrand, Marianne, Dolly Chugh and Sendhil Mullainathan (2005), Implicit Discrimination, *American Economic Review Papers and Proceedings*, Vol. 95(2), pp. 94-8.
- [12] Blank, Rebecca M. (1991), The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from *The American Economic Review*, *American Economic Review*, Vol. 81(5), pp. 1041-67.
- [13] Bolger, N. and T. Kellaghan (1990), Method of Measurement and Gender Differences in Scholastic Achievement, *Journal of Educational Measurement*, Vol. 27, pp. 165-74.
- [14] Bon Reis, Susan, I. Phillip Young and James C. Jury (1999), Female Administrators: A Crack in the Glass Ceiling, *Journal of Personnel Evaluation in Education*, Vol. 13(1), pp.71-82.
- [15] Borjas, George J. and Matthew S. Goldberg (1978), Biased Screening and Discrimination in the Labor Market, *American Economic Review*, Vol. 68(5), pp. 918-922.

- [16] Broder, Ivy E. (1993), Review of NSF Economics Proposals: Gender and Institutional Patterns, *American Economic Review*, Vol. 83(4), pp. 964-70.
- [17] Byrne, D. (1971), *The Attraction Paradigm* (New York: Academic Press)
- [18] Chan, Nixon and Peter E. Kennedy (2002), Are Multiple-Choice Exams Easier for Economics Students? A Comparison of Multiple Choice and Equivalent Constructed Response Exam Questions, *Southern Economic Journal*, Vol. 68, pp. 957-71.
- [19] Chattopadhyay, Raghavendra and Esther Duflo (2004), Women as Policy Makers: Evidence from a Randomized Experiment, *Econometrica*, Vol. 72(5), pp.1409-43.
- [20] Cole, Stephen, Leonard Rubin and Jonathan Cole (1978), *Peer Review at the National Science Foundation* (Washington, DC: National Academy of Sciences)
- [21] Cornell, Bradford and Ivo Welch (1996), Culture, Information, and Screening Discrimination, *Journal of Political Economy*, Vol. 104(3), pp. 542-71.
- [22] Dee, Thomas S. (2004), Teachers, Race, and Student Achievement in a Randomized Experiment, *The Review of Economics and Statistics*, Vol. 86(1), pp. 195-210.
- [23] Dillingham, Alan E., Marianne E. Ferber and Daniel S. Hamermesh (1994), Gender Discrimination by Gender: Voting in a Professional Society, *American Economic Review*, Vol. 47(4), pp. 622-33.
- [24] Ely, Robin J. (1994), The Effects of Organizational Demographics and Social Identity on Relationships among Professional Women, *Administrative Science Quarterly*, Vol. 39(2), pp. 203-38.
- [25] Ely, Robin J. (1995), The Power in Demography: Women's Social Constructions of Gender Identity at Work, *Academy of Management Journal*, Vol. 38(3), pp. 589-634.
- [26] Fryer, Roland G. and Glenn Loury (2005), Affirmative Action and Its Mythology, *Journal of Economics Perspectives*, Vol. 19(3)
- [27] Gallois, C., V. J. Callan, and J. Palmer (1992), The Influence of Applicant Communication Style and Interviewer Characteristics on Hiring Decisions, *Journal of Applied Social Psychology*, Vol. 22, pp. 1041-60.
- [28] Gneezy, Uri, Muriel Niederle and Aldo Rustichini (2003), Performance in Competitive Environments: Gender Differences, *Quarterly Journal of Economics*, Vol. 118(3), pp. 1049-74.
- [29] Goldberg, Caren B. (2005), Relational Demography and Similarity Attraction in Interview Assessments and Subsequent Offer Decisions, *Group and Organization Marketing*, Vol. 30(6), pp. 597-624.
- [30] Goldin, Claudia and Cecilia Rouse (2000), Orchestrating Impartiality: The Effect of "Blind" Auditions on Female Musicians, *American Economic Review*, Vol. 90(4), pp. 715-41.

- [31] Graves, L. M. and G. N. Powell (1995), The Effect of Sex Similarity on Recruiters' Evaluations of Actual Applicants: A Test of the Similarity-Attraction Paradigm, *Personnel Psychology*, Vol. 48(1), pp. 85-98.
- [32] Graves, L. M. and G. N. Powell (1996), Sex Similarity, Quality of the Employment Interview and Recruiters' Evaluations of Actual Applicants, *Journal of Occupational and Organizational Psychology*, Vol. 69, pp. 243-61.
- [33] Greene, B. (1997), Verbal Abilities, Gender and the Introductory Economics Course: A New Look at Old Assumptions, *Journal of Economic Education*, Vol. 28(Winter), pp. 13-30.
- [34] Grupo Especial de Trabajo sobre Buen Gobierno (2006), *Proyecto de Código Unificado de Recomendaciones de Buen Gobierno de Sociedades Cotizadas*
- [35] Hamermesh, Daniel and Jeff E. Biddle (1994), Beauty and the Labor Market, *American Economic Review*, Vol. 84(5), pp. 1174-94.
- [36] Hamermesh, Daniel and Amy M. Parker (2005), Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity, *Economics of Education Review*, Vol. 24(4), pp. 369-76.
- [37] Heilman, M. E., Martell, R. F., and M.C. Simon (1988), The Vagaries of Sex Bias: Conditions Regulating the Undervaluation, Equivaluation and Overvaluation of Female Job Applicants, *Organizational Behavior and Human Decision Processes*, Vol. 41, pp. 98-110.
- [38] Instituto Nacional de Estadística, Estadística de la Enseñanza Universitaria, several issues.
- [39] Instituto Nacional de Estadística, Estadística de la Enseñanza Superior, several issues.
- [40] Lavy, Victor (2008), Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment, *Journal of Public Economics*, forthcoming.
- [41] Lazear, Edward P. and Sherwin Rosen (1990), Male-Female Wage Differentials in Job Ladders, *Journal of Labour Economics*, Vol. 8(1)
- [42] Lee, V. E., J. B. Smith, and M. Cioci (1993). Teachers and Principals: Gender-Related Perceptions of Leadership and Power in Secondary Schools, *Educational Evaluation and Policy Analysis*, Vol. 15(2), pp. 153-80.
- [43] Lumsden, Keith G. and Alex Scott (1987), The Economics Student Re-Examined: Male-Female Differences in Comprehension, *Journal of Economic Education*, Vol.18(4), pp. 365-75.
- [44] Ministerio de la Presidencia, *Boletín Oficial del Estado*, selected issues (available at <http://www.boe.es>)
- [45] Niederle, Muriel and Lise Vesterlund (2007), Do Women Shy away from Competition? Do Men Compete too Much?, *Quarterly Journal of Economics*, Vol. 122(3). pp. 1067-1101.
- [46] Palacios-Huerta, Ignacio and Oscar Volij (2008), Experientia Docet: Professionals Play Minimax in Laboratory Experiments, *Econometrica*, Vol. 76(1), pp. 71-115.

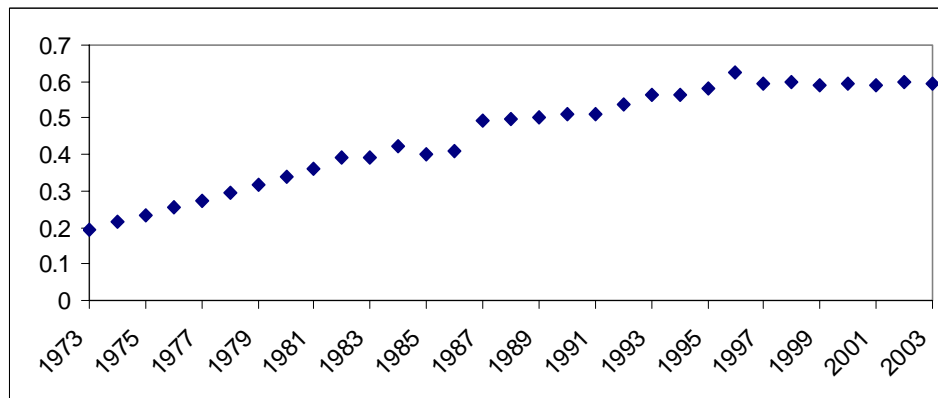
- [47] Pande, Rohini (2003), Can Mandated Political Representation Provide Disadvantaged Minorities Policy Influence? Theory and Evidence from India, *American Economic Review*, Vol. 93(4), pp. 1132-151.
- [48] Papke, L. E. and J. Wooldridge (1996), Econometric methods for fractional response variables with an application to 401(k) plan participation rates, *Journal of Applied Econometrics*, Vol. 11, pp. 619–32.
- [49] Phelps, Edmund (1972), The Statistical Theory of Racism and Sexism, *American Economic Review*, Vol. 62, pp. 659-61.
- [50] Price, Joseph and Justin Wolfers (2007), Racial Discrimination among NBA Referees, NBER Working Paper 13206
- [51] Quintero, Esther (2008), How are Job Applicants Disadvantaged by Gender Based Double Standards in a Natural Setting, unpublished dissertation, Cornell University
- [52] Steele, Claude M. (1997), A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance, *American Psychologist*, Vol. 52(6), pp. 613-29.
- [53] Walstad, William B. and William H. Becker (1994), Achievement Differences on Multiple-Choice and Essay Tests in Economics, *The American Economic Review*, Vol. 84(2), Papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economic Association, pp. 193-96.
- [54] Wiley, M. G. and A. Eskilson (1985), Speech Style, Gender Stereotypes and Corporate Success: What if Women Talked more like Men?, *Sex Roles*, Vol. 12, pp. 993-1007.
- [55] Wooldridge, Jeffrey M. (2003), Cluster-Sample Methods in Applied Econometrics, *American Economic Review Papers and Proceedings*, Vol. 93(2), pp. 133-8.

Figure 1 Female Corps Members (share), 1970-2004



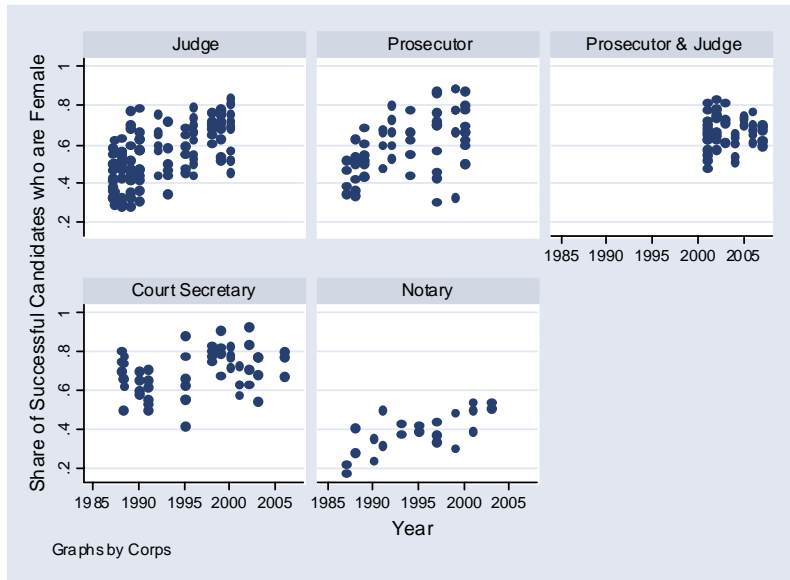
Source: Authors' calculations based on official Corps rankings.

Figure 2 Female Law Graduates in Spain (share), 1973-2004



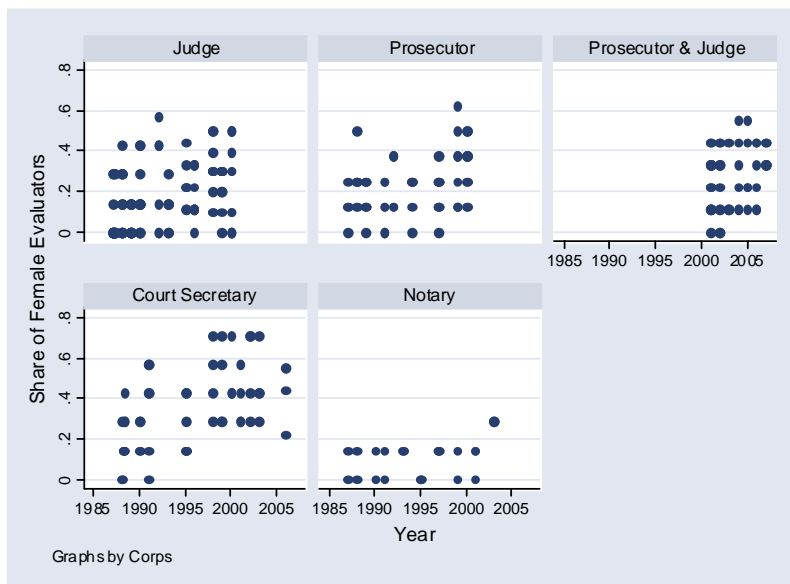
Source: Estadística de la Enseñanza Universitaria en España and Estadística de la Enseñanza de la Superior en España, Instituto Nacional de Estadística, several years.

Figure 3 Female Successful Candidates (share), by type of examination and committee, 1987-2007



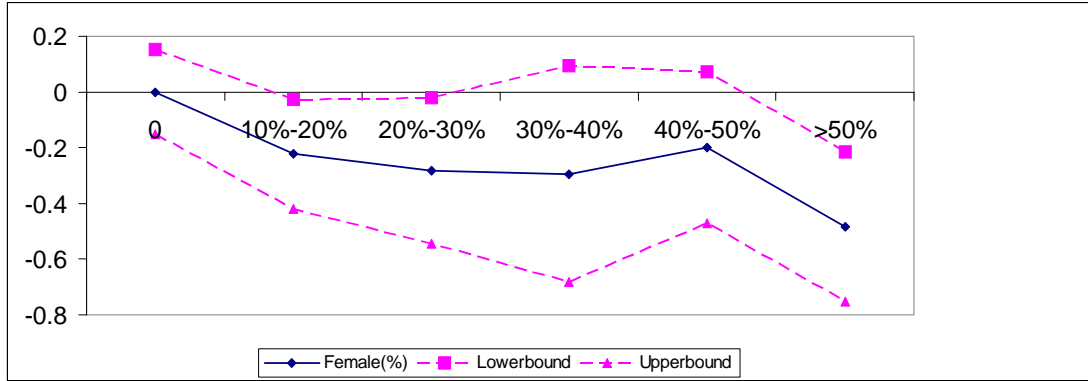
Source: Authors' calculations.

Figure 4 Female Committee Members (share), by type of examination and committee, 1987-2007



Source: Authors' calculations.

Figure 5 Gender Composition of Committees and Probability of Success: Nonlinearities (Table 7)



Source - Estimates in column (6) in Table 7, where the dependent variable is the proportion of successful candidates who are female and this variable is normalized equal to zero for committees with no female members. The lowerbound (upperbound) represents the estimated coefficient minus (plus) two standard deviations.

Table 1: Rules and Format of Exams, by Type of Examination

| | Court Secretary (1) | Judge (2) | Judge and Prosecutor (3) | Notary (4) | Prosecutor (5) |
|--------------------------|--|-------------------------------------|--|----------------------------------|-------------------------------------|
| Number of members | 7/9 ¹ | 7/9/10 ² | 9 | 7 | 8 |
| President | Judge ³ | Superior Court Judge | Superior Court Judge or Prosecutor ⁴ | Registrar or Notary President | Superior Court Prosecutor |
| Other members | | | | | |
| Court secretary | 2/3 | 0/1 | 1 | 0 | 0 |
| Judge | 1/0 | 3/4 | 2 | 0 | 1 |
| Lawyer | 1 | 1 | 1 | 0 | 1 |
| Notary | 0 | 0 | 0 | 2/1 | 0 |
| Professor | 1 | 2 | 1 | 1 | 1 |
| Prosecutor | 0/1 | 1 | 2 | 0 | 3 |
| Registrar | 0 | 0 | 0 | 1/2 | 0 |
| State lawyer | 1/0 | 1 | 1 | 1 | 1 |
| Number of stages | 2 | 2 | 2/3 | 4 | 2 |
| Exam format ⁵ | Written(1987-95) Oral(1998-2003) Multiple Choice and Oral(2006) | Written(1987-95) Oral(1996-2000) | Oral(2001-2002) Multiple Choice and Oral(2003-2007) | Oral and Written | Written(1987-94) Oral(1997-2000) |

Notes: 1,2/Over the years, the rules regarding the composition of some committees have slightly changed. For more details on committee composition rules see the appendix. 3/From a Superior Court of Justice until 1997. 4/Every odd (even) committee is presided by a judge (prosecutor), alternatively. 5/Written exams have to be read by candidates in front of committees.

Table 2: Available Exams and Committees, by Type of Examination

| | Total | Court Secretary | Judge | Judge and Prosecutor | Notary | Prosecutor |
|---------------------------------------|-------|---|---|---------------------------------------|--|--|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Available exams ¹ | | 1988,1990,1991, 1995,1998,1999, 2000,2001,2002, 2003,2006 | 1987,1988,1989, 1990,1992,1993, 1995,1996,1998, 1999,2000 | 2001,2002, 2003,2004, 2005,2006, 2007 | 1987,1988,1990, 1991,1993,1995, 1997,1999,2001, 2003 | 1987,1988,1989, 1991,1992,1994, 1997,1999,2000 |
| Number of exams | 51 | 13 | 12 | 7 | 10 | 9 |
| Written/Oral ² | 20/21 | 6/7 | 8/4 | 0/7 | - | 6/3 |
| Not all positions assigned | 25 | 9 | 6 | 6 | 0 | 4 |
| Average number of committees per exam | 6.1 | 4.2 | 9.8 | 7.1 | 2.2 | 7.2 |
| Number of committees | 309 | 54 | 118 | 56 | 22 | 65 |

Notes: 1/Usually there is only one exam per year, but in 1988 there were three exams for court secretary positions. In 1987 there were two exams for judge positions. 2/Notary exams include both written and oral stages.

Table 3: Descriptive Statistics - Characteristics of Successful Candidates at the Individual Level, by Type of Examination

| | Total | Court Secretary | Judge | Judge and Prosecutor | Notary | Prosecutor |
|------------------------|----------------|-----------------|----------------|----------------------|----------------|----------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Female (%) | 58.6 | 69.3 | 57.0 | 66.6 | 38.1 | 58.0 |
| Exam grade | 0.62 (0.16) | 0.63 (0.16) | 0.56 (0.17) | 0.66 (0.11) | 0.71 (0.11) | 0.59 (0.16) |
| Number of observations | 7700 | 1708 | 2359 | 1442 | 1070 | 1121 |

Notes: standard deviations in parentheses. Grades are normalized between zero and one.

Table 4: Descriptive Statistics - Committee Level

| | Mean (1) | Standard deviation (2) | Minimum (3) | Maximum (4) |
|--------------------------------------|-------------|---------------------------|----------------|----------------|
| Successful candidates | | | | |
| Positions available | 26.84 | 15.57 | 8 | 108 |
| Positions assigned | 24.92 | 13.99 | 8 | 92 |
| Female | 14.76 | 9.09 | 3 | 71 |
| Male | 10.16 | 7.77 | 1 | 50 |
| Female (%) | 60.17 | 15.2 | 18.03 | 92.86 |
| Evaluators | | | | |
| Number of committee members | 7.98 | 1.06 | 7 | 10 |
| Female share in the committee | 0.22 | 0.17 | 0 | 0.71 |
| Female share=0 | 0.20 | 0.41 | 0 | 1 |
| $0.1 \leq \text{Female share} < 0.2$ | 0.29 | 0.45 | 0 | 1 |
| $0.2 \leq \text{Female share} < 0.3$ | 0.19 | 0.39 | 0 | 1 |
| $0.3 \leq \text{Female share} < 0.4$ | 0.13 | 0.34 | 0 | 1 |
| $0.4 \leq \text{Female share} < 0.5$ | 0.14 | 0.34 | 0 | 1 |
| Female share > 0.5 | 0.05 | 0.22 | 0 | 1 |
| Age | 45.88 | 5.18 | 29.25 | 60.86 |
| Experience | 0.62 | 0.59 | 0 | 2.56 |
| Ranking | 0.63 | 0.12 | 0.19 | 0.92 |

Notes: Given that committees have between seven and ten members, in no case there was a female share between 0 and 0.1. In nine cases the share of female in the committee was equal to 0.5. These cases have been classified taking into consideration the gender of the president (whose vote prevails in case of tie).

Table 5: Descriptive Statistics - Candidates to Judge and Prosecutor (2003-2007) and Court Secretary (2006) Positions

| | All (1) | Female (2) | Male (3) |
|-------------------------------|------------|---------------|-------------|
| Turns out | 83.66 | 84.33 | 82.13 |
| Pass stage one | 42.11 | 41.14 | 44.34 |
| Pass stage two | 10.49 | 10.34 | 10.82 |
| Pass stage three ¹ | 5.06 | 4.94 | 5.34 |
| Experience | | | |
| First time | 23.02 | 23.36 | 22.22 |
| Second time | 16.01 | 16.21 | 15.54 |
| Third time | 13.65 | 13.70 | 13.55 |
| Fourth time | 47.32 | 46.73 | 48.69 |
| Number of candidates | 24530 | 17099 | 7431 |

Notes: figures represent percentages. All three stages are qualifying. The first stage is a multiple choice test, the second and third stages are oral exams. 1/In the period considered, all candidates who passed the third stage obtained a position, except for one candidate to Judge and Prosecutor positions in 2004.

Table 6: Descriptive Statistics - Committee Members, by Profession

| | Total | | Court Secretary | | Lawyer | | Judge | |
|------------------------|----------------|----------------|-----------------|----------------|----------------------|----------------|----------------|----------------|
| | Female (1) | Male (2) | Female (3) | Male (4) | Female (5) | Male (6) | Female (7) | Male (8) |
| Share | 0.23 | 0.77 | 0.43 | 0.57 | 0.33 | 0.67 | 0.20 | 0.80 |
| Age | 40.1 (0.46) | 48.1 (0.34) | 46.6 (0.86) | 52.1 (0.94) | . | . | 43.0 (0.71) | 51.8 (0.51) |
| Ranking | 0.64 (0.01) | 0.62 (0.01) | 0.71 (0.03) | 0.68 (0.03) | . | . | 0.57 (0.02) | 0.67 (0.01) |
| Experience | 0.58 (0.04) | 0.67 (0.02) | 0.70 (0.11) | 0.89 (0.10) | 1.02 (0.12) | 0.60 (0.06) | 0.58 (0.07) | 0.62 (0.04) |
| Number of observations | 2467 | | 222 | | 287 | | 625 | |
| | Professor | | Prosecutor | | Registrar and Notary | | State Lawyer | |
| | Female | Male | Female | Male | Female | Male | Female | Male |
| Share | 0.16 | 0.84 | 0.21 | 0.79 | 0.06 | 0.94 | 0.19 | 0.81 |
| Age | . | . | 36.7 (0.59) | 49.3 (0.60) | 36.8 (2.85) | 47.1 (1.11) | 29.4 (0.41) | 36.9 (0.70) |
| Ranking | . | . | 0.66 (0.02) | 0.62 (0.01) | 0.73 (0.09) | 0.63 (0.03) | 0.61 (0.04) | 0.45 (0.02) |
| Experience | 0.52 (0.11) | 0.68 (0.05) | 0.37 (0.07) | 0.90 (0.05) | 0 (0) | 0.01 (0.01) | 0.16 (0.06) | 0.56 (0.05) |
| Number of observations | 427 | | 530 | | 84 | | 289 | |

Notes: standard errors in parentheses. The ranking index has been calculated based on the member's ranking within her age group. A higher number means the member is higher in the ranking. Lawyers and professors are not Judicial Corps members hence we cannot observe their age nor rank. All committee members come from these eight Corps, except for three civil servants of type A from the Ministry of Justice.

Table 7: Gender Composition of Committee and Probability of Success

| | Dependent variable: successful candidates | | | | | | | |
|----------------------------|---|---------|----------|---------|------------|----------|-----------|---------|
| | Log female | | Log male | | Female (%) | | Log total | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Female share | -0.22** | | 0.31* | | -0.52** | | -0.01 | |
| | (0.10) | | (0.15) | | (0.21) | | (0.06) | |
| 0.10 ≤ Female share < 0.20 | | -0.08 | | 0.16** | | -0.24** | | 0.03 |
| | | (0.06) | | (0.06) | | (0.11) | | (0.03) |
| 0.20 ≤ Female share < 0.30 | | -0.12 | | 0.17** | | -0.28** | | 0.00 |
| | | (0.08) | | (0.07) | | (0.13) | | (0.03) |
| 0.30 ≤ Female share < 0.40 | | -0.12 | | 0.17 | | -0.30 | | -0.00 |
| | | (0.08) | | (0.14) | | (0.19) | | (0.05) |
| 0.40 ≤ Female share < 0.50 | | -0.09 | | 0.11 | | -0.20 | | 0.01 |
| | | (0.06) | | (0.10) | | (0.13) | | (0.04) |
| Female share > 0.5 | | -0.18** | | 0.33*** | | -0.51*** | | 0.01 |
| | | (0.08) | | (0.10) | | (0.13) | | (0.06) |
| Log positions | 0.60*** | 0.55*** | 0.81 | 0.94 | -0.22 | -0.39 | 0.65*** | 0.68*** |
| | (0.13) | (0.13) | (0.53) | (0.56) | (0.41) | (0.47) | (0.24) | (0.25) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes | yes | yes | yes | yes |
| Adjusted R ² | 0.772 | 0.771 | 0.749 | 0.751 | 0.475 | 0.479 | 0.905 | 0.904 |
| Number of observations | 309 | 309 | 309 | 309 | 309 | 309 | 309 | 309 |

Note: standard errors clustered at the exam level in parentheses. The control group is no women in committees

The dependent variable in columns (5) and (6) is $\log(\text{proportion female}/(1-\text{proportion female}))$.

*significant at 10%; **significant at 5%; ***significant at 1%.

Table 8: Gender Composition of Committee and Probability of Success, By Decade

| | Dependent variable: successful candidates | | | |
|-------------------------------|---|-------------------|--------------------|------------------|
| | Log female (1) | Log male (2) | Female (%) (3) | Log total (4) |
| 1987-2007 | | | | |
| Minority of female evaluators | -0.10 (0.06) | 0.16*** (0.06) | -0.25** (0.10) | 0.02 (0.03) |
| Majority of female evaluators | -0.17** (0.08) | 0.34*** (0.09) | -0.51*** (0.13) | 0.02 (0.06) |
| F test (p-value) | 0.21 | 0.02 | 0.01 | |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.773 | 0.754 | 0.484 | 0.905 |
| Number of observations | 309 | 309 | 309 | 309 |
| 1987-1996 | | | | |
| Minority of female evaluators | -0.05 (0.08) | 0.14** (0.06) | -0.19 (0.12) | 0.04 (0.04) |
| Majority of female evaluators | -0.20** (0.09) | 0.29*** (0.04) | -0.49*** (0.11) | -0.01 (0.05) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.687 | 0.760 | 0.381 | 0.879 |
| Number of observations | 159 | 159 | 159 | 159 |
| 1997-2007 | | | | |
| Minority of female evaluators | -0.20*** (0.05) | 0.23 (0.13) | -0.43*** (0.14) | -0.03 (0.05) |
| Majority of female evaluators | -0.26*** (0.09) | 0.42** (0.16) | -0.69*** (0.19) | -0.02 (0.08) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.829 | 0.717 | 0.288 | 0.929 |
| Number of observations | 150 | 150 | 150 | 150 |

Notes: standard errors clustered at the exam level in parentheses. All regressions control for the logarithm of the number of positions. The null hypothesis for the F test is that the minority coefficient and the majority coefficient are the same. The dependent variable in column (3) is $\log(\text{proportion female}/(1-\text{proportion female}))$. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 9: Gender Composition of Committee and Probability of Success, by Type of Examination

| Examination | Dependent variable: successful candidates | | | |
|-------------------------------|---|-------------------|--------------------|-------------------|
| | Log female (1) | Log male (2) | Female (%) (3) | Log total (4) |
| Court secretary | | | | |
| Minority of female evaluators | -0.13*** (0.02) | 0.13 (0.10) | -0.26** (0.11) | -0.04** (0.02) |
| Majority of female evaluators | -0.24*** (0.07) | 0.36** (0.14) | -0.60*** (0.15) | -0.05 (0.06) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.883 | 0.634 | 0.243 | 0.920 |
| Number of observations | 54 | 54 | 54 | 54 |
| Judge | | | | |
| Minority of female evaluators | -0.08 (0.09) | 0.20** (0.09) | -0.28* (0.15) | 0.03 (0.04) |
| Majority of female evaluators | -0.08 (0.05) | 0.32*** (0.05) | -0.40*** (0.09) | 0.06** (0.03) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.475 | 0.316 | 0.396 | 0.435 |
| Number of observations | 118 | 118 | 118 | 118 |
| Judge and Prosecutor | | | | |
| Minority of female evaluators | -0.13** (0.04) | 0.09 (0.30) | -0.22 (0.26) | -0.05 (0.12) |
| Majority of female evaluators | -0.18** (0.06) | 0.22 (0.34) | -0.41 (0.32) | -0.02 (0.13) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.819 | 0.667 | 0.127 | 0.878 |
| Number of observations | 50 | 50 | 50 | 50 |
| Notary | | | | |
| Minority of female evaluators | -0.17 (0.17) | 0.11 (0.09) | -0.28 (0.25) | . |
| Majority of female evaluators | . | . | . | . |
| Exam dummies (Corps*Year) | yes | yes | yes | |
| Adjusted R ² | 0.688 | 0.913 | 0.580 | |
| Number of observations | 22 | 22 | 22 | |
| Prosecutor | | | | |
| Minority of female evaluators | -0.05 (0.12) | 0.10 (0.09) | -0.15 (0.13) | 0.03 (0.08) |
| Majority of female evaluators | 0.03 (0.12) | 0.12 (0.09) | -0.08 (0.13) | 0.03 (0.08) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.638 | 0.748 | 0.214 | 0.965 |
| Number of observations | 65 | 65 | 65 | 65 |

Notes: standard errors clustered at the exam level in parentheses. All regressions control for the logarithm of the number of positions. There is no variation in the number of successful candidates across committees for notaries. The dependent variable in column (3) is $\log(\text{proportion female}/(1-\text{proportion female}))$. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 10: Committee Characteristics and Probability of Success

| | Dependent variable: successful candidates | | | |
|---------------------------------|---|-------------------|--------------------|-------------------|
| | Log female (1) | Log male (2) | Female (%) (3) | Log total (4) |
| Minority of female evaluators | -0.09 (0.06) | 0.16*** (0.06) | -0.26** (0.10) | 0.02 (0.03) |
| Majority of female evaluators | -0.17** (0.08) | 0.36*** (0.08) | -0.52*** (0.12) | 0.02 (0.06) |
| Age of committee members | 0.00 (0.01) | -0.00 (0.01) | 0.00 (0.01) | 0.00 (0.00) |
| Experience of committee members | 0.00 (0.05) | -0.04 (0.07) | 0.04 (0.11) | -0.01 (0.03) |
| Ranking of committee members | 0.16 (0.12) | -0.50** (0.20) | 0.67** (0.28) | -0.06 (0.07) |
| Log positions | 0.56*** (0.15) | 0.88 (0.52) | -0.32 (0.43) | 0.66*** (0.25) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.772 | 0.758 | 0.490 | 0.904 |
| Number of observations | 309 | 309 | 309 | 309 |

Note: standard errors clustered at the exam level in parentheses. Age and ranking are available for all committee members except for professors and private sector lawyers. ranking equals one (zero) if the individual ranks first (last) among Corps members born the same year. The control group is no women in committees. The dependent variable in column (3) is $\log(\text{proportion female}/(1-\text{proportion female}))$. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 11: Evaluation or Performance? Written vs. Oral Exams

| | Dependent variable: successful candidates | | | | | | | |
|---------------------------------|---|-------------------|------------------|-------------------|-------------------|--------------------|-------------------|-----------------|
| | Log female | | Log male | | Female (%) | | Log total | |
| | Oral (1) | Written (2) | Oral (3) | Written (4) | Oral (5) | Written (6) | Oral (7) | Written (8) |
| Minority of female evaluators | -0.09 (0.14) | -0.11 (0.08) | 0.23 (0.14) | 0.14* (0.07) | -0.32 (0.19) | -0.25 (0.15) | 0.03 (0.10) | 0.00 (0.02) |
| Majority of female evaluators | -0.11 (0.14) | -0.29** (0.11) | 0.44** (0.16) | 0.28*** (0.06) | -0.55** (0.20) | -0.57*** (0.17) | 0.07 (0.11) | -0.07 (0.05) |
| Age of committee members | 0.02* (0.01) | -0.00 (0.01) | -0.00 (0.01) | -0.00 (0.01) | 0.02 (0.01) | 0.00 (0.01) | 0.01 (0.01) | -0.00 (0.00) |
| Experience of committee members | 0.03 (0.06) | -0.18** (0.09) | -0.06 (0.09) | 0.00 (0.09) | 0.09 (0.13) | -0.18 (0.14) | 0.00 (0.03) | -0.09 (0.06) |
| Ranking of committee members | 0.02 (0.18) | 0.27 (0.20) | -0.66 (0.41) | -0.47 (0.28) | 0.68 (0.53) | 0.73 (0.43) | -0.11 (0.13) | -0.02 (0.11) |
| Log positions | 0.73*** (0.14) | 0.93 (1.24) | 0.81 (0.59) | 2.95 (2.77) | -0.08 (0.50) | -2.02 (3.63) | 0.75*** (0.25) | 1.78* (0.95) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes | yes | yes | yes | yes |
| Adjusted R ² | 0.808 | 0.707 | 0.638 | 0.642 | 0.118 | 0.303 | 0.908 | 0.857 |
| Number of observations | 150 | 137 | 150 | 137 | 150 | 137 | 150 | 137 |

Note: standard errors clustered at the exam level in parentheses. Age and ranking is available for all members except for professors and private sector lawyers. Ranking equals one (zero) if the individual ranks first (last) among individuals born the same year. The control group is no women in committees. Notary exams have not been included because they are half written half oral. The dependent variable in columns (5) and (6) is $\log(\text{proportion female}/(1-\text{proportion female}))$. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 12: Who is Discriminated? Evidence from Exams where not all Positions were Assigned

| | Dependent variable: successful candidates | | | |
|---------------------------------|---|------------------|-------------------|------------------|
| | Sample: exams where not all positions were assigned | | | |
| | Log female (1) | Log male (2) | Female (%) (3) | Log total (4) |
| Minority of female evaluators | -0.04 (0.10) | 0.14 (0.12) | -0.18 (0.19) | 0.02 (0.04) |
| Majority of female evaluators | -0.10 (0.12) | 0.32** (0.13) | -0.43** (0.20) | 0.03 (0.07) |
| Age of committee members | 0.01 (0.00) | 0.01 (0.01) | 0.00 (0.01) | 0.01** (0.00) |
| Experience of committee members | -0.04 (0.06) | -0.04 (0.08) | -0.01 (0.12) | -0.03 (0.04) |
| Ranking of committee members | 0.05 (0.16) | -0.33 (0.29) | 0.38 (0.38) | -0.03 (0.13) |
| Log positions | 0.42** (0.15) | 0.46 (0.45) | -0.04 (0.40) | 0.46** (0.21) |
| Exam dummies (Corps*Year) | yes | yes | yes | yes |
| Adjusted R ² | 0.800 | 0.669 | 0.427 | 0.887 |
| Number of observations | 166 | 166 | 166 | 166 |

Note: standard errors clustered at the exam level in parentheses. Age, experience and ranking are available for all members except for professors and private sector lawyers. Ranking equals one (zero) if the individual ranks first (last) among individuals born the same year. The dependent variable in column (3) is $\log(\text{proportion female}/(1-\text{proportion female}))$. The control group is no women in committees. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 13: Who is Discriminated? Evidence from Grades

| | Dependent variable: Exam Grade | |
|---------------------------------|--------------------------------|------------------|
| | Male (1) | Female (2) |
| | Minority of female evaluators | 0.02** (0.01) |
| Majority of female evaluators | 0.05* (0.03) | 0.00 (0.02) |
| Age of committee members | -0.000 (0.001) | 0.001 (0.001) |
| Experience of committee members | -0.01 (0.01) | -0.00 (0.01) |
| Ranking of committee members | -0.10*** (0.04) | 0.02 (0.03) |
| Exam dummies (Corps*Year) | yes | yes |
| Adjusted R ² | 0.472 | 0.391 |
| Number of observations | 2145 | 3337 |

Note: standard errors clustered at the committee level in parentheses. Grades are normalized between zero and one. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 14: Is the Allocation of Candidates to Committees Random?

| | Dependent variable: | | |
|--|---------------------|------------------|----------------|
| | Female [Probit] | Exam Grade | |
| | (1) | Male (2) | Female (3) |
| $0.20 \leq \text{Female share} < 0.30$ | -0.01 (0.02) | -0.46 (0.47) | 0.17 (0.32) |
| $0.30 \leq \text{Female share} < 0.40$ | 0.02 (0.02) | -0.31 (0.53) | 0.15 (0.35) |
| $0.40 \leq \text{Female share} < 0.50$ | 0.00 (0.01) | -0.75* (0.40) | 0.15 (0.27) |
| Female committee share > 0.5 | 0.01 (0.02) | -0.38 (0.57) | 0.07 (0.37) |
| Exam dummies (Corps*Year) | yes | yes | yes |
| Adjusted R ² | 0.003 | 0.549 | 0.578 |
| Predicted probability | 0.68 | . | . |
| Number of observations | 11309 | 3295 | 7034 |

Note: standard errors in parentheses. The control group is no women in committees. *significant at 10%;

significant at 5%; *significant at 1%.

Table 15: Probability of Success by Gender and Gender Composition of Committee, 2003-2007

| | Candidates who passed stage one (multiple choice test) | |
|--|--|---|
| | Dependent variable: | |
| | Exam Grade [Tobit] (1) | Probability to be hired [Probit] (2) |
| Female (=1 if female candidate) | 1.53 (1.08) | 0.01 (0.01) |
| $0.20 \leq \text{Female share} < 0.30$ | -0.75 (1.43) | -0.01 (0.01) |
| Female* $0.20 \leq \text{Female share} < 0.30$ | -0.14 (1.66) | -0.00 (0.01) |
| $0.30 \leq \text{Female share} < 0.40$ | 1.57 (1.59) | 0.01 (0.02) |
| Female* $0.30 \leq \text{Female share} < 0.40$ | -1.43 (1.76) | -0.01 (0.01) |
| $0.40 \leq \text{Female share} < 0.50$ | 1.32 (1.29) | 0.00 (0.01) |
| Female* $0.40 \leq \text{Female share} < 0.50$ | -1.32 (1.52) | -0.01 (0.01) |
| Female committee share > 0.5 | 3.09** (1.54) | 0.03* (0.02) |
| Female*Female committee share > 0.5 | -3.86** (1.79) | -0.03** (0.01) |
| Multiple choice mark | 0.88*** (0.04) | 0.01*** (0.00) |
| One year of experience | 0.86 (0.97) | 0.01 (0.01) |
| Two years of experience | 0.52 (1.01) | 0.01 (0.01) |
| Three years of experience | -2.31** (0.92) | -0.01 (0.01) |
| Year dummies | yes | yes |
| Adjusted/Pseudo R ² | 0.08 | 0.146 |
| Predicted probability | . | 0.07 |
| Number of observations | 10329 | 10329 |

Notes: standard errors in parentheses. The exams consist of three qualifying stages: (i) a multiple choice test, (ii) an oral test, (iii) an oral test. Here we consider candidates who passed the multiple choice test and hence assigned to committees for the oral evaluation. The passing grade is 25 and the maximum is 50. The control group is male candidates evaluated by committees with one woman ($0.10 \leq \text{Female share} < 0.20$). The tobit in column (1) is left-censored at 25 as we only observe grades for those passing the exam. *significant at 10%; **significant at 5%; ***significant at 1%.