I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12530

Selecting Talent: Gender Differences in Participation and Success in Competitive Selection Processes

Lídia Farré
Francesc Ortega

AUGUST 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12530

Selecting Talent: Gender Differences in Participation and Success in Competitive Selection Processes

Lídia Farré
University of Barcelona, IAE-CSIC and IZA

Francesc Ortega
CUNY, Queens College and IZA

AUGUST 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Selecting Talent: Gender Differences in Participation and Success in Competitive Selection Processes*

We investigate whether competitive selection processes generate gender inequality in the context of a prestigious graduate fellowship program. All applications are scored remotely by expert reviewers and the highest ranked are invited to an in-person interview. The data show a very large gender gap in success rates: women's success rate is 36% lower than men's. About one third of this gap is due to the lower grades of female candidates, which is surprising given women's higher GPA in the population of college graduates. Adjusting for GPA and a rich set of fixed-effects, women's success rate remains 16% lower than for comparable male candidates. We show that this gap is explained by reviewers engaging in gender balancing. Namely, reviewers favor the minority gender in each field of study but, except for STEM, all fields are female-dominated. Our simulations show that the interview plays an important role, but the quantitative scoring has a more profound effect on the award allocation. Merging administrative records on the population of graduates from a large university, we document an important gender gap in participation. We find that high-GPA female graduates are much less likely to apply to the fellowship program. The combination of the gender gaps in participation and success in the program imply that high-GPA female graduates are almost 50% less likely to obtain a fellowship than their male counterparts.

JEL Classification: J3, J7

Keywords: glass ceiling, gender, education

Corresponding author:
Francesc Ortega
Department of Economics
Queens College, CUNY
300A Powdermaker Hall
65-30 Kissena Blvd.
Queens, New York 11367
USA
E-mail: fortega@qc.cuny.edu

* We are grateful to the *La Caixa Foundation* and the *University of Barcelona* for providing access to their internal data and procedures, and for their overall support. We are particularly indebted to David Casado and Ampar Asensi. We also thank Lola Collado, Julen Esteban-Pretel, Ada Ferrer, Albrecht Glitz, Amy Hsin, Libertad Gonzalez, Pedro Rey, Nuria Rodriguez-Planas, Kevin Shih, Todd Sorensen and Suleyman Taspinar for their advice. We also thank seminar participants at UB-IEB, CUNY and Universitat Pompeu Fabra.

1 Introduction

Despite large gains in recent decades, a substantial wage gender gap persists in many countries. Many studies have linked this to the under-representation of women in high-earnings, high-status occupations, often referred to as the *glass ceiling*, recently reviewed in Bertrand (2018).

A large body of literature has found that a myriad of factors contribute to the existence of the gender *glass ceiling*. Women choose lower earning degrees and remain under-represented in STEM (Bertrand et al. (2010a)), have a higher demand for flexible schedules (Goldin (2014)), underperform under pressure, and actively avoid competitive settings (Gneezy et al. (2003), Iriberri and Rey-Biel (2019b)). But other factors also play important roles, including the existence of certain of gender norms (Fernández et al. (2004), Bertrand et al. (2015), Bertrand and Duflo (2017)), child penalties (Kleven et al. (2019)), and taste-based or statistical discrimination (Bertrand and Duflo (2017)).

Our paper focuses on a different explanation that has not received as much attention in the literature. Namely, the structure of the selection processes that provide access to top positions in the labor market may stack cards against female candidates. Access to entry-level positions for highly skilled workers often entails a two-stage selection process. First, applications are often distilled into a few quantitative scores that are used to sift through the pool of applications. The highest-ranked applications are then invited to an interview, which will determine who is selected to fill the position. With minor modifications, the same type of selection process is used in many contexts, ranging from admission to graduate or undergraduate programs to recruiting in the private sector and academia.

There are reasons to believe that this type of selection process creates gender inequality. For instance, several studies have argued that women often underperform in competitive environments (Gneezy et al. (2003), Iriberri and Rey-Biel (2019b)). This may also be the case during high-stakes interviews, which would imply that the selection process outlined above penalizes female candidates. Clearly, if the positions under consideration do not require performing under high pressure, the female penalty arising from this type of selection process introduces an inefficiency. Similarly, some recent studies have pointed out that quantitative scoring may disproportionately penalize women or minorities for reasons unrelated to candidate quality. Rivera and Tilcsik (2019) show that seemingly irrelevant aspects, such as the range of the scale used in the scoring of applications, can introduce gender gaps because of gender stereotypes of brilliance. Kolev et al. (2019) have argued that language differences in written statements also generate gender gaps in the success rates for obtaining funding for research projects. As a result of these biases, award allocations may be distorted and women may be penalized. As far as we can tell, no studies have analyzed whether this type of two-stage selection process does in fact engender gender inequality and, if so, which component of the process is responsible. Furthermore, women may self-select out of this type of selection process in order to avoid a highly competitive environment, or due to a perception that the employers/reviewers discriminate against them. If this is the case, the reduced female participation may compound with gender inequalities resulting from the selection process and result in a larger overall gender penalty. Identifying whether aspects of recruiting and selection processes generate gender inequality can provide useful insights into how these processes can be improved.

We address these questions using a new dataset with unique information on a highly competitive selection process with the two-stage structure outlined above. Specifically, we use detailed data on a large number of applicants competing to obtain a highly prestigious fellowship to carry out graduate studies: Spain's *La Caixa* Fellowship Program. The main purpose of these fellowships is to provide funding for Spanish citizens to conduct graduate (Master's or Ph.D) studies abroad in any field of study through three sub-programs defined by the geographic location of the destination universities.[1]

Within Spain, this fellowship program is widely known and considered very prestigious. As a result, it is also highly competitive and fewer than 9% of all complete applications are funded. A recent study by Garcia-Montalvo (2014) showed that the stakes are high. The labor market careers of individuals who were awarded the fellowship experienced a large and persistent boost, relative to candidates that were highly ranked but were not selected.

The selection process is structured in two stages. First, all applications are remotely evaluated by two randomly-assigned expert reviewers who provide quantitative scores along several pre-established dimensions. The top ranked applications are short-listed and advance to an in-person panel interview. We obtained detailed information on all applicants to the *La Caixa* graduate fellowship program over the period 2014-2018, containing demographic and academic information, along with the quantitative scores used at the screening stage, and the outcome of the panel interview. These data allow us

[1] Some fellows that have gone on to successful academic careers in Economics are Jordi Gali (1984), Xavier Sala-i-Martin (1984), Luis Garicano (1992) or, more recently, Marti Mestieri (2005) and Eduardo Morales (2005).

to analyze the determinants of success within the program. Furthermore, we obtained administrative records on all graduates from a large Spanish university, the *University of Barcelona*, for the period 2009-2018. The two datasets were linked at the individual level, preserving anonymity, allowing us to extend the analysis by examining the decision to participate in the fellowship program

Our analysis yields several interesting findings. First of all, we document the existence of a very large raw gap in success rates between male and female candidates: the success rate for females is 3.8 percentage points lower (42% of the mean success rate).[2] Accounting for differences in university of origin, field of study and other fixed-effects lowers the gender gap modestly (to 3.3 percentage points). Heterogeneity in GPA plays a larger role: adjusting for individual differences in GPA, the gender penalty in success rates falls to 1.5 percentage points (17% of the mean success rate) within the same field of study, university of origin and sub-program. The reduction in the conditional gender gap when controlling for GPA indicates that female applicants have lower grades, on average, than male applicants. This is a surprising finding, given that female college graduates earn higher grades than men, suggesting a high degree of self-selection along gender-GPA lines in the decision to apply to the fellowship program.

Second, our analysis of the quantitative scores produced during the remote evaluation shows that reviewers play an important role in generating gender differences among candidates who are otherwise similar in terms of GPA and along other dimensions. The data show that female candidates systematically receive lower scores than male candidates with the same GPA along each of the three dimensions considered (Transcripts & CV, Proposal and Letters).

Third, we document a large degree of heterogeneity in the size of the success gender gaps across fields of study. After adjusting for GPA, female candidates in Health & Life Sciences are 5.1 percentage points *less* likely to obtain a fellowship than their male counterparts. The picture is similar in Social Sciences and in Arts & Humanities, though the female penalty is smaller in the latter. In contrast, female candidates in STEM are 1.4 percentage points *more* likely to obtain a fellowship than comparable male candidates. Thus, it appears that reviewers engage in *gender balancing*, favoring the minority gender within their respective fields. As a result, female candidates experience a gender premium in male-dominated STEM fields. However, because women are the majority gender in all other fields, they suffer an overall penalty. Accordingly, the largest female penalty is found in Health & Life Sciences, the field with the largest

[2] The raw (unconditional) success rate for women is 36% lower than for men.

proportion of females.

Fourth, to identify the role played by the different aspects of the selection process we simulate counterfactual award allocations on the basis of different criteria. The findings show that the selection process, and particularly the remote scoring of applications, profoundly influences the allocation of awards within fields of study, though the aggregate effects are muted by countervailing effects across fields due to *gender balancing*. In STEM the selection process *increases* the share of females among winners by 10 percentage points relative to an allocation based purely on GPA. In contrast, in female-dominated fields the selection process lowers the winners' female share by more than 5 percentage points. We also show that, while the interview plays a noticeable role in generating gender inequality, quantitative scoring at the screening stage plays the larger role in most fields. In particular, we find that reviewers' scores of candidates' *Transcripts & CV* increase gender inequality uniformly across all fields, mirroring differences in GPA. Additionally, the scores assigned to *Letters of reference* appear to reinforce the pattern of *gender balancing*, and we fail to find any systematic effect of the scores regarding the quality of the *Proposal*.

Last, we analyze the decision to apply to the fellowship program using our matched dataset. Our main finding is that high-GPA women are much less likely to participate than male graduates with the same GPA in the same field of study. According to our analysis, the combined gender gaps in participation and in success within the program lead to a situation where high-GPA female graduates are close to 50% less likely than their male peers to obtain a *La Caixa* fellowship.

Our work is related to the rapidly evolving literature on the the factors driving gender gaps in the labor market.[3] The most relevant studies in the context of our paper are those focusing on high-pay, highly skilled occupations. Bertrand et al. (2010b) and Azmat and Ferrer (2017) study gender gaps among MBAs and lawyers, respectively. In both cases they find that the earnings gap between men and women are driven by differences in career interruptions and working hours, often tied to childbearing, and to gender differences in career aspirations. In a recent study, Boustan and Langan (2019) have documented a variety of factors that account for the severe under-representation of women in Economics departments. Our paper contributes to this literature by analyzing the role of the selection process itself in generating gender inequality in outcomes.

Until recently, much less attention has been given to gender differences in participation. Carpio and Guadalupe (2019) document gender differences in the decision to enter

[3] A recent review on this literature can be found at Bertrand (2018).

the technology sector driven by social norms and show that de-biasing messages can be an effective policy tool to increase female participation. Hospido et al. (2019) examine gender differences in career progression in central banking and find that women are less likely to apply for promotion than men. However, conditional on applying, women are more likely to be approved. Our study also contributes to this literature by analyzing the decision by college graduates on whether to apply to the graduate fellowship program. In contrast to the previous studies, our analysis is based on a much larger dataset, containing all graduates from a large university over an 8-year period.

Our work is also related to the studies on gender differences in performance in competitive settings. Several studies have found evidence of female underperformance under high pressure in experimental settings (Gneezy et al. (2003), Iriberri and Rey-Biel (2017) and Iriberri and Rey-Biel (2019b)) and in real-world settings (Azmat et al. (2016) and Montolio and Taberner (2018)).[4] An important manifestation of these differences is that women try to avoid highly competitive environments (Niederle and Vesterlund (2007)), which ties in with our analysis of the role of gender as a factor determining participation in the fellowship program. Our paper also analyzes the roles of the remote evaluation and the panel interview in generating gender inequality. Clearly, candidates' performance during the interview may be affected by the perception and response to a high-stakes environment. Thus, our findings will be relevant to understand whether women's performance is negatively affected, relative to men, in this particular context.

Our work is also related to the literature studying to what degree the design of the tools used to judge merit affect the measurement of gender gaps, with a particular emphasis on the role of reviewers. Rivera and Tilcsik (2019) show that quantitative performance ratings of faculty teaching evaluations can also generate gender inequality. More specifically, they find that the range of the scale used affects the measured gender gap because of gender stereotypes of brilliance. Kolev et al. (2019) argue that written proposals can also lead to gender differences unrelated to quality. These authors analyzed data on grant proposals competing for funding and found that female-authored proposals received lower scores due to differences in writing stye. Specifically, they found that women used narrower (topic-specific) words to describe their contribution, while men used broader language. In both cases, reviewers' choices led to inefficient allocations. Several other studies have zoomed into the role of reviewers in generating gender

[4] Azmat and Petrongolo (2014) provide a review of the experimental literature in regards to gender differences in labor market outcomes and discuss the strengths and limitations in terms of actual workplace settings.

inequality. Card et al. (2019) document gender differences in peer-review evaluations in Economics journals, showing that reviewers (regardless of their gender) set a higher bar for female-authored papers. Our dataset contains individualized data on the scores submitted by each reviewer on each application, providing a window to examine the role played by reviewers on candidates' success across different fields of study.

Last, some studies have focused on the effects of the gender of reviewers on outcomes. In the context of a national competition for judge positions in Spain, Bagues and Esteve-Volart (2010) show that the number of female evaluators in the committee negatively affects the female share among successful candidates, suggesting that female-majority committees over-estimate the quality of male candidates. A later study by Bagues et al. (2017) using data on national evaluations to obtain tenured professor positions in Spain and in Italy produced similar findings: a higher number of women in the evaluation committee increases neither the quality nor quantity of selected females. Our data contains information on the gender of reviewers, allowing us to investigate the presence of interactions between the gender of reviewers and candidates.

The remainder of our paper is structured as follows. Section 2 presents our data sources. Section 3 discusses our econometric specification. Section 4 presents our estimates of the gender gaps in success rates. Section 5 turns to the estimation of gender gaps in reviewer scores. Section 6 examines heterogeneity across fields of study. Section 7 summarizes the results from our simulations. Section 8 analyzes the determinants of participation in the fellowship program, and Section 9 concludes.

2 Data

Our analysis is based on two sources of data: detailed information on all applicants to the graduate fellowship program funded and administered by the *La Caixa Foundation* (LCF for short) over the period 2014-2018, and administrative records for all graduates of the *University of Barcelona* (UB) between 2009 and 2018. The two datasets were linked at the individual level, preserving anonymity, which allow us to analyze both the decision to participate in the program and the determinants of success, conditional on participation.

2.1 The LCF Applicants Dataset

The LCF is a private financial institution in Spain that has been providing graduate fellowships since 1982. To date, the LCF has funded more than 4,500 awards, totaling over 220 million euros in funding. Our data contains applications to three separate sub-programs, defined by the geographic location of the destination universities. Roughly speaking, half of the applications in our data seek funding for studies in European countries (other than Spain), one quarter aim at studying in North American or Asian universities, and the remaining quarter seek funding for conducting doctoral studies in Spanish institutions.[5]

The program has grown over time and, currently, over 1,800 applications are received annually, resulting in about 130 fellowships per year. Our data covers the period 2014-2018 and contains complete information on roughly 8,100 applicants with a female share of 55% and an average success rate of 8.8%.[6]

At the time of submitting the application, candidates self-select into 30 narrow fields of study. From this point on, the applications go through a two-stage selection process. In stage 1, every application is randomly assigned to two reviewers who are experts in the narrow field selected by the applicant. Reviewers score applications along three dimensions: Transcripts & CV, Quality of the proposal, and Letters of reference. At the time of scoring applications, reviewers have access to the whole application package, including full transcripts. An overall score is computed for each application and a ranking is produced on the basis of this score. In our data, about 19% of the applications go on to the second stage, which consists of an in-person interview by a 5-person panel of experts. The panel is asked to focus on the quality of the proposal of study as well as the potential of the candidate, and roughly half of those interviewed are awarded the fellowship. The data contain information on individual characteristics such as age, gender and grade point average (GPA), in addition to the scores assigned by reviewers to each application.

Table 1 presents some descriptive statistics. According to our data, the success rate in the first stage of the selection process is about 19%, and the overall success rate (considering both stages) is around 9%. The average age is 27 and only 30% of the applicants are age 30 or older. The distribution of applicants across fields of study

[5] About 46% of the awards funded studies in European countries (other than Spain), 30% in Spain and 24% in North America or Asia. The Spain program only funds doctoral degrees, whereas the Europe and North America/Asia programs fund Master's, Ph.Ds, and other graduate degrees.

[6] We drop from the analysis roughly 500 applications pertaining to candidates that obtained their undergraduate degrees outside of Spain due to differences in the grading system.

is quite balanced: 30% STEM, 26% Social Sciences, and 22% each for Health & Life Sciences and Arts & Humanities. The table also summarizes the scores of the remote evaluation, on a 1-8 scale, which average around 6.4 along each of the three dimensions evaluated by the reviewers.

It is also interesting to compare the GPA distributions of the candidates to the fellowship program. Figure 1 plots the distributions by gender of the applicant. The Figure shows a larger mass of male candidates at the top of the grade distribution, compared to female candidates. As shown in Figure 2, this pattern is present in all fields of study, but more striking in STEM and Health & Life Sciences. This fact is striking when we take into account that among recent university graduates in Spain, as in many other countries, women graduate with higher GPA than men.

2.2 The UB Graduates Dataset

We obtained administrative data for all students that graduated from the *University of Barcelona* (UB) between 2009 and 2018. The data contains 56,946 graduates and includes socio-economic and academic information: age, gender, graduation year, field of study, parental education and GPA.

We merged these data with the LCF applicants dataset, preserving anonymity. The data show that, during our period of analysis, 432 students from the UB applied to the fellowship program, which amounts to a 0.76% participation rate. This rate increases substantially among high-GPA students, the relevant population interested in graduate studies and with more realistic chances of winning the award. Among students with GPA above the 75th percentile, the participation rate raises to 1.22%, and it climbs to 1.70% among students with GPA above the 90th percentile.

We also use these data to characterize the GPA distributions of male and female graduates. Figure 3 pools graduates of all fields of study and shows that the distribution for female graduates is clearly shifted to the right, relative to males, implying that female graduates have better grades. The contrast between this figure and the corresponding one referring to the pool of applicants to the LCF program (Figure 1) is rather striking. Disaggregating by field of study, Figure 4 shows that the female GPA distribution is also shifted to the right, relative to men's, in Social Sciences and Health & Life Sciences. In Arts & Humanities, there are practically no differences between the GPA distributions of male and female candidates. The picture is more nuanced in STEM, where women's grades are clearly better in the bottom half of the grade distribution, but somewhat worse in the top half. These gender differences across fields in GPA will play an important role in our analysis in the following sections.

3 Econometric specification

3.1 Success rates

Our first goal is to estimate the gender gap in success rates conditional on GPA and other individual characteristics. To do so we consider a model where the dependent variable is an indicator variable y_i, taking a value of one if individual i is awarded the fellowship:

$$y_i = \alpha + \beta Fem_i + X_i'\delta + \varepsilon_i, \tag{1}$$

where Fem_i is a dummy variable indicating if candidate i is female. Characteristics vector X_i includes the GPA of the candidate, age, and a rich set of fixed-effects for year of application, program/destination, field of study and university of origin. We refer to β in Equation (1) as the *conditional* gender gap in success rates.

3.2 Reviewer scores

Our data contains information on the scores assigned by each of the two reviewers that evaluate each application. Because each reviewer assigns scores to many applications, we are able to account for reviewer heterogeneity through fixed-effects. Specifically, we consider different versions of the following model:

$$Score_{i,r} = \alpha_r + \beta Fem_i + \lambda Fem_i \times FemRev_r + \delta X_i + \varepsilon_{i,r}, \tag{2}$$

where $Score_{i,r}$ is the score received by candidate i from reviewer r, α_r is a reviewer fixed-effect and Fem_i is a dummy variable for the gender of the applicant. Indicator variable $FemRev_r$ takes a value of one when the reviewer is female. As before, vector X_i includes applicant characteristics, such as GPA and age, and a rich set of fixed-effects (year, program, field of study and university of origin).

This specification is very flexible and implies the following marginal effects: coefficient β identifies the female-male differential arising from *male reviewers* and $\beta + \lambda$ identifies the gender differential arising from *female reviewers*. Thus, λ identifies whether male and female reviewers penalize female candidates (over male candidates) to a different degree.

We cluster standard errors by application since it is likely that the assessments of different reviewers are correlated for the same application. For instance, reviewers's assessment of a candidate's reference letters are rlikely to impact their assessments of a candidate on all dimensions.

4 Gender gaps in success rates

Our first goal is to estimate the determinants of success in the program, as in Equation (1), with an emphasis in investigating if there exists a gender gap in success rates after conditioning on observable characteristics.[7]

Table 2 presents our findings. The top panel of the table reports the gender differential in success rates, relative to men. The first column shows a raw female penalty of 3.82 percentage points. This is a very large gap, as it amounts to 43% of the mean success rate (8.83%) in our data.

First, we investigate to what extent this gap is due to gender differences in year of application, sub-program/destination, field of study and university of origin.[8] Column 2 controls for age of the candidate, by including an indicator for applicants age 30 or older, which corresponds to the 75th percentile in the age distribution, and also includes year fixed-effects. Older candidates have a success rate that is 2.59 percentage-points lower, indicating that these candidates had more difficulty in completing their undergraduate degrees. At any rate, the gender gap remains practically unchanged compared to column 1. Column 3 adds program fixed-effects, which lowers the female differential to 3.63 percentage points (a 5.5% reduction relative to column 2). Column 4 includes fixed-effects for field of study, reducing slightly the gender differential by an additional 5%, relative to column 3. Last, column 5 adds university of origin fixed-effects, which lowers the gender differential by 4% relative to column 4. In this specification we still observe a female penalty of 3.31 percentage points, or 37% of the mean success rate. These estimates show that women are disproportionately concentrated in fields of study and programs with lower success rates. However, these factors account by less than 15%

[7]We define a candidate as *successful* if he or she was awarded a fellowship. We note that some successful candidates declined awards. This is a rare event but it happens occasionally, for instance when a candidate has won a similar fellowship from another funding agency.

[8]We consider four fields of study: STEM, Health and Life Sciences, Arts and Humanities, and Social Sciences. The three fellowship programs are North America/Asia, Europe or Spain (Ph.D. only).

(0.5 percentage-points) of the raw gender gap in success rates. Thus, the gender gap in success rates is due to individual differences within the same program, field of study and university of origin.

Next, we ask to what extent this large gender gap in success rates can be explained by differences in the academic quality of the candidates. The bottom panel of Table 2 addresses this question by controlling for GPA.[9] More specifically, to allow for non-linearities, we include dummy variables for GPA in the range 50-75 percentile, 75-90 percentile, and above the 90th percentile. Focusing on column 5, which includes the whole set of fixed-effects, we observe a very large impact of GPA on the success rate in the program. The success rate for candidates with a GPA above the 90th percentile is about 30 percentage points higher than for candidates with GPA below the median. Thus, academic performance, as measured by GPA, is the most important determinant of success in the selection process.

The second important observation is that the female penalty in success rates drops importantly when accounting for GPA, indicating that female candidates have lower GPA than male candidates. As discussed earlier, Figure 1 shows a larger mass of male candidates at the top of the grade distribution, compared to female candidates. This pattern is present in all fields of study, but more striking in STEM and Health & Life Sciences (Figure 2). The reversal of the ranking of GPA distributions by gender, as compared to the population of college graduates, stems from the fact that (i) only students with strong academic records are interested in pursuing graduate studies, and (ii) among high-GPA students, females are less likely to participate in a competitive fellowship program aimed primarily at pursuing those studies abroad. We will provide a detailed analysis of the participation decision later in the paper.

Lastly, the estimates in column 5 (bottom panel), show that the *unexplained* gender gap in success rates is 1.5 percentage-points (or 17% of the mean success rate) among candidates with the same GPA (and age) within the same program, field of study and university of origin. The goal of the rest of the section is to investigate further the nature of the unexplained gender gap in success gap rates. One specific, but important, question is whether the gap arises at the interview stage, or rather during the remote evaluation. A simple, initial approach to this question is to include the reviewer scores

[9] Our measure of GPA is self-reported and there is experimental evidence documenting gender differences in aversion to lying (Croson and Gneezy (2009) and Childs (2012)). However, applicants to the program also submit an official transcript, which unfortunately is not part of our data. This transcript is available to reviewers, severely limiting the incentive to misreport.

during the remote evaluation as control variables.[10] If the gender gap vanishes we can then conclude that the gender gap stems from the panel interview.

Table 3 presents the results. As can be seen in column 5, reviewer scores are highly significant determinants of the success in the program, with *Transcripts & CV* playing the largest role. This is consistent with the manner in which reviewer scores are aggregated to decide which candidates advance to the second stage of the selection process.[11] Secondly, the gender gap in success rates now decreases to 1.09 percentage points, a 27% reduction relative to the 1.50 gap reported in column 5 of Table 2. Thus, gender differences in scores play a role in shaping the outcomes of the selection process, but the panel interview may also contribute to generating gender inequality in success rates. The next section scrutinizes further the reviewer scores to determine whether gender differences are present and, if so, in which of the three dimensions of the reviews.

5 Gender gaps in quantitative scores

Gender differences in performance in competitive settings have received a great deal of attention in the literature (e.g. Iriberri and Rey-Biel (2019a)). In comparison, much less attention has been devoted to examining whether gender differences arise also in the quantitative scoring of applications. While often regarded as gender neutral, two recent studies call this assumption into question. Kolev et al. (2019) analyzed blinded-review scores of grant proposals and found that female candidates received lower scores than men that could not be explained by applicant quality, reviewer characteristics or topic. The authors concluded that reviewers were swayed by gender differences in the language used in the written proposals. Similarly, Rivera and Tilcsik (2019) documented gender differences in quantitative performance ratings in the context of faculty teaching evaluations also unrelated to quality differences, but purely due to the numeric scale used to rate teachers.

Our data contains reviewer-level scores on each application. We now build a dataset where observations are defined at the applicant-reviewer level and estimate models that include reviewer fixed-effects, as in Equation (2). These fixed-effects will absorb all reviewer-specific characteristics that affect equally all candidates, such as the severity of

[10] Recall that each application receives scores by two reviewers along three dimensions: (i) Transcripts & CV, (ii) Quality of the proposal, and (iii) Reference letters. The scoring scale ranges from 1 to 8.

[11] The administrators of the LCF fellowship program compute a weighted average of the three scores given by $Score = 0.50 Transcripts\&CV + 0.3 Proposal + 0.2 Letters$.

each individual reviewer. However, the fixed-effects will not absorb differences arising from reviewer gender bias, that is, treating differently applications on the basis of gender (or some other criterion).

It is helpful to consider first a simpler specification, where we do not make a distinction on the basis of the gender of the reviewer. The first three columns in Table 4 present these results. As before, the top panel provides estimates of the gender gaps that do not control for GPA. The table shows that female candidates receive lower scores, relative to male candidates from the same university of origin, in the same field of study and applying to the same fellowship program in the same year. The largest gap is observed in the category of *Transcripts & CV*, and amounts to 0.16 points (on a 1-8 scale). Similarly, women also obtain lower scores in the quality of their proposal and regarding the reference letters, but the differences are smaller along these dimensions. The larger gap in *Transcripts & CV* is not surprising, given the gender differences in GPA documented earlier.

Let us now turn to the bottom panel, where we control for GPA. Once again, we find evidence of gender gaps along the three dimensions scored by reviewers, though the size of the gaps is now smaller and the differences across fields have also shrank down. Relative to the mean score, the gaps are fairly small, in the range of 0.6% to 1% of the mean scores. However, their accumulated effect amounts to a 0.52 percentage-point gap in the probability of success, or 35% of the unexplained 1.50 gap.[12] In sum, the evidence suggests that female candidates receive (slightly) lower scores during the remote evaluation than observationally equivalent males. This is the case in terms of the reviewers' assessment of candidates' academic ability ($Transcripts\&CV$), which could be due to gender bias on the part of the reviewers, but also to differences in the transcripts of male and female applicants beyond their average grade.[13] We also find that reviewers give lower scores to female candidates in regards to letters of reference and quality of the proposal, consistent with the findings in Kolev et al. (2019).

Next, we ask whether the *gender of the reviewer* plays a role in explaining the gender differences in scores. To do so we turn to columns 4-6 in Table 4, which display the estimates for the model in Equation (2). The results suggest that both male and female

[12] To quantify the accumulated effects on the success rate of the gender differences in stage-1 scores, we proceed as follows. We average the estimated gender differences for each item, multiply them by respective marginal effect on the success rate (obtained in Table 3), and add up across items: $-(5.4 \times 0.06) - (2.7 \times 0.04) - (2.2 \times 0.04) = -0.52$.

[13] For instance, it could be the case that male candidates have higher grades in courses considered *tougher* even though females are more regular and graduate with a higher overall GPA.

reviewers penalize *female* candidates similarly in terms of *Transcripts & CV*. However, *female* reviewers appear to give lower scores to *female* candidates regarding the quality of their *Reference letters* and *Proposal of study*, consistent with the findings in Bagues and Esteve-Volart (2010) and Bagues et al. (2017). It is worth noting that the latter two dimensions are more subjective aspects of the assessment, and therefore more prone to reflect stereotypes or other types of reviewer bias.

6 Differences by field of study: gender balancing

In Spain, as in many other countries, women account for the majority of enrollment in colleges and universities. However, the female presence in some fields is much larger than in others. This pattern is also present among applicants to the LCF program. Ranking the academic fields on the basis of the female share among applicants shows that STEM is the field with the lowest female share (34%), followed by Social Sciences (60%), Arts & Humanities (63%) and Health & Life Sciences (68%). In comparison, the female share among all applicants to the fellowship program is 55%. Thus, except for STEM, all fields exhibit above-average shares of female candidates.

These differences across fields of study suggest it may be interesting to conduct the analysis separately by field. The results are collected in Table 5. The dependent variable is an indicator for success in the program, as before. The top panel illustrates the existence of unadjusted female penalties across the four fields of study. However, there exists large heterogeneity across fields, ranging from a small female penalty of 1.4 percentage points in STEM to a large 7.7 percentage-point penalty in Health & Life Sciences. This heterogeneity is not explained by cross-field differences in gender gaps in GPA. As we discussed earlier, the grade distribution for female applicants is shifted to the left relative to the male distribution *both* in STEM and Health & Life Sciences (Figure 2). Nonetheless the gender gaps in success rates between the two fields are very different. Furthermore, after adjusting for individual GPA (in the bottom panel), a large degree of heterogeneity across fields remains, ranging from a female *advantage* of 1.4 percentage points in STEM to a 5.1 percentage-point *penalty* in success rates in Health & Life Sciences. The estimates also indicate small penalties (of 1.2 percentage points) in Social Sciences and Arts & Humanities.

This pattern suggests that reviewers engage in *gender balancing*. That is, female candidates appear to be treated more favorably than comparable men in male-dominated fields, but they are penalized in female-dominated fields. Unfortunately for them, women are the majority among applicants in all fields of study, with the sole exception of STEM. *Gender balancing* is more striking when we compare the most male-dominated field (STEM) with the most female-dominated field (Health & Life Sciences). In STEM the female share among applicants is 34% while in Health & Life Sciences it is 68%. However, in STEM the female share among winners is 6 percentage points lower than the female share among candidates. In contrast, the female share among winners in Health & Life Sciences is 17 percentage-points lower than the female share among applicants in that field.[14]

The applicant-reviewer dataset containing the scores of the remote evaluation provides further evidence of *gender balancing*. As shown in Table 6, female candidates receive systematically lower scores than comparable males in all female-dominated fields. For instance, in Health & Life Sciences, Arts & Humanities, and Social Sciences, women's scores are close to 1% lower than for their male counterparts along the three dimensions considered. In contrast, in STEM we find no significant gender difference in the score for Transcripts & CV, but a positive and significant *advantage* for women, relative to comparable men, in the quality of the *Proposal* and *Reference Letters* in the range of 1.5% to 2% of the mean score. Once again, unexplained gender differences in scores appear in the most subjective dimensions of the application.

In sum, the analysis in this section provides clear evidence that the selection process is not gender neutral when we focus on the selection of winners within each field of study and that, to some extent, the action takes place during the remote evaluation of applications. Specifically, we observe that reviewers favor the candidates that are under-represented in terms of gender among the pool of applicants in each field of study. Because of these countervailing effects across fields, analyzing all fields pooled together severely understates the role of the selection process in shaping outcomes along gender lines.

[14]Further evidence to support this interpretation is obtained from the estimation of a model where the dependent variable is an indicator for success in the program and, besides the usual set of regressors, we now include an interaction between the gender of the candidate and the share of females in his/her field of study. The estimates show that there is no evidence of a gender penalty in success in fields with low female shares. But a success penalty arises when the share of female candidates in the field of study increases. This Table is available upon request.

7 Simulations

We complement the analysis of the previous sections with a number of simulation exercises that will help determine which specific aspects of the selection process engender gender inequality. In these simulations we compare alternative award allocations varying the selection criteria. Across the different scenarios we keep constant the number of fellowships actually awarded in each year, program and field of study.[15]

Our starting point is a scenario where awards are distributed purely on the basis of GPA (simulation 1).[16] In this case, the award allocation will be based purely on information that existed prior to the selection process. Thus the selection process would play no role, and any gender differences will be driven by differences in the GPA distributions for male and female candidates.[17]

Another important scenario we consider allocates awards to the candidates attaining the highest scores during the remote evaluation. Specifically, we aggregate the three preselection scores into a single numerical measure, which is then used to select the winners (simulation 3). Importantly, the comparison between the share of female winners (or the gender gap in success rates) in simulations 1 and 3 provides a measure of the gender differences introduced by the remote evaluation. In turn, comparison of the allocation of awards between simulation 3 and the actual allocation is a reflection of the gender differences arising from the interview. We also consider a few additional scenarios, which will allow us to decompose the roles of the three quantitative scores: *Transcripts & CV*, quality of the *Proposal*, and *Letters of reference*.

7.1 Overall selection process

One of the main questions we would like to answer is whether the LCF selection process generates gender inequality. In other words, does the process mitigate or exacerbate

[15]In each of the years in our data about 140 fellowships were offered to successful candidates. Because some offers were declined by the candidates, the actual number of awards was around 130 per year.

[16]In some occasions, we encountered candidates within a field of study, program and year, with the same exact GPA. In those cases we break the tie using the score for the letters of reference and, if necessary, the score for *Transcripts & CV*. Alternatively, we could have broken the ties through random number generation. The advantage of specifying a tie-breaking rule is that the results can then be reproduced more easily. As far as we can tell, the main results do not depend on the tie-breaking rule adopted.

[17]LCF requires that at least half of the fellowships in any given year be allocated to the fields of STEM and Health & Life Sciences combined. However, this restriction does not appear to have much influence on the resulting allocation. The data show that these two fields account for 52% of the applications and receive 58% of the awards.

the initial gender differences in the pool of applicants? Our baseline is a scenario where awards are allocated solely on the basis of candidates' GPA, keeping the total number of awards by year, program and field of study as in the data. Comparing this allocation to the actual one in the data we will be able to tease out whether the selection process generates gender inequality.

Let us first consider the results for all fields pooled together, collected in the top panel of Table 7. Column 1 summarizes the actual award allocation. The top panel displays the results for all fields pooled together. Among the 8,099 applications received, 54.8% were female. The selection process resulted in 715 candidates receiving offers for a fellowship. Among the winners, the fraction of women was 44.1%, 10.7 percentage-points lower than the female share among candidates, and their average score in the remote evaluation was 7.4. As shown in column 2, had awards been allocated purely on the basis of GPA, the female share among winners would have been only 0.9 percentage-points lower (43.2%). Hence, in net terms, the selection process actually *reduces* gender inequality slightly. Thus the gender inequality in success rates observed in the data is mostly a reflection of the large gender differences in GPA between male and female candidates documented earlier, which profoundly shape the allocation of awards. It is also worth noting that the overlap between the sets of winners in the actual and counterfactual GPA-based allocations is small: only 37.3% of the would-be winners on the basis of GPA were actual winners. Thus the the selection process dramatically influences the set of winners, even though this effect is largely neutral in terms of gender.[18]

Informed by our earlier findings, we suspect that the small aggregate effect of the selection process on gender inequality may be masking heterogeneous effects across fields of study. As we show next, this is indeed the case. The selection process profoundly influences the allocation of awards within fields of study, but with countervailing effects across fields. Once again, we find a clear pattern of *gender balancing*. This is seen most clearly in the first column of Table 8, which reports the difference between the female share among winners in the *actual* allocation and the *counterfactual* GPA-based allocation. In male-dominated fields (STEM), the selection process increases the share of

[18] Table 12 reports an analogous table but measuring gender inequality as the difference between the success rates of men and women. None of the main conclusions drawn changes when using this different metric. The award allocation based solely on GPA features a very large gender gap in success rates: 11.1% of male candidates were selected, compared to 7.0% of women. Thus the female success rate was 37% lower than the male rate, indicating that female applicants have a much lower GPA than male applicants on average. In the actual allocation the success rates of males and females are 10.9% and 7.1%, respectively. This amounts to a 35% gender gap. Once again, in net terms, the selection process reduces inequality slightly relative to a purely GPA-based allocation.

females among winners by 10 percentage points, relative to the GPA-based allocation. In contrast, in female-dominated fields the selection process lowers the female share among winners. More specifically, the selection process lowers the winner female share by 7.2 and 5.2 percentage points in Health & Life Sciences and in Social Sciences, respectively. The field of Arts & Humanities is a bit exceptional: despite being moderately female-dominated, the selection process is roughly gender neutral.

An important implication of our findings is that the effect of the selection process on gender inequality is highly heterogeneous across fields of study, with a clear pattern of *gender balancing* with offsetting effects across fields of study. It is also worth noting that some of the inequality-inducing mechanisms emphasized in the literature, such as female underperformance under high pressure, or systematic gender differences in reviewer scores of candidates' proposals do not appear to play a large role in our specific context. Thus these mechanisms may be less universal than previously thought and can be mitigated when selection processes are carefully designed and implemented.

7.2 Quantitative scoring vs. in-person interview

Several studies have argued that the main factor explaining female gaps in the labor market is underperformance of women in high-stakes environments (e.g. Iriberri and Rey-Biel (2017)). In our context, if this hypothesis is true, we would expect the success gender gap to arise at the panel interview (stage 2). Furthermore, to the extent that underperformance is linked to the ways in which girls are raised, as opposed to boys, one would expect to find evidence across all fields of study. More recently, researchers have also examined the potential for gender bias to arise from quantitative scoring of applications. For instance, Kolev et al. (2019) found evidence of gender bias in competitive grant applications, arising from gender differences in writing style. These studies motivate the question of which of the two parts of the LCF selection process plays a larger role in generating gender inequality in success rates: the initial remote evaluation of applications, or the in-person, panel interview?

To decompose the overall effect of the selection process into the contribution of each of the two stages, we compare three scenarios: the GPA-based allocation (simulation 1), an award allocation based solely on the quantitative scores generated in the remote evaluation (simulation 3), and the actual award allocation, which takes into account the outcomes of the interview, the quantitative scores, and candidates' GPA. These scenarios allow us to decompose the overall effect of the selection process into the effect of the remote evaluation (quantitative scoring) and the interview.

Table 8 summarizes the results of the decomposition, which are based on the more detailed statistics reported in Table 7. The first column simply reports the overall effect of the selection process discussed earlier. For instance, in STEM the actual allocation increases the female share among winners by 10 percentage points, relative to the counterfactual GPA-based allocation.

The second column isolates the role of the *interview* by comparing a counterfactual allocation based solely on the (combined) scores produced during the remote evaluation to the actual allocation, determined by both the remote evaluation and the panel interview. The results show that the interview hardly affected the female share among winners in STEM. In the other fields, the interview did play a substantial role by reducing the winners female share in Health & Life Sciences by 3.6 percentage-points, while increasing it by 3.9 and 2.1 points in Social Sciences and Arts & Humanities, respectively. In sum, the interview shaped in important ways the gender balance among winners in all fields but STEM. However, while the interview appears to have penalized women in some fields, it favored them in others.

Next, we turn to the role of the *quantitative scoring* of applications during the remote evaluation. To isolate its effect we compare the outcomes in the scenarios where fellowships are awarded purely based on GPA (simulation 1) to the case where the allocation is based on the scores produced during the remote evaluation. Column 3 in Table 8 summarizes the results. Clearly, the remote evaluation played the key role in shaping the large positive effect on the winners' female share in STEM. In all other fields the quantitative scoring had a negative effect on women's success. Quantitatively, these effects were as large (in absolute value) as those of the interview in Health & Life Sciences, and much larger in Social Sciences (and in STEM). Only in Arts & Humanities was the effect of the remote evaluation slightly lower than that of the panel interview. Furthermore, column 3 clearly illustrates the pattern of *gender balancing* documented earlier: the remote evaluation is largely responsible for increasing the success rates of the candidates whose gender is in the minority in each field of study.

One way to compare the contributions of the two stages of the interview is to sum the absolute values of the effects of each stage across fields. As shown at the bottom of the table, the remote evaluation appears to have had a more profound effect in shaping our outcome of interest. This is perhaps not surprising, given that the remote evaluation influences the outcome of all applications, whereas the panel interview only affects the minority (roughly 20%) that reach the second stage of the selection process. Our results underscore the findings in Rivera and Tilcsik (2019) and Kolev et al. (2019) by showing that quantitative scoring of applications is not gender neutral due to reviewers' behavior. However, our findings emphasize the heterogeneity of these effects across fields of study.

7.3 Components of the scoring: credentials, proposal and letters of reference

Kolev et al. (2019) analyzed the success rates of male and female applicants seeking funding for research projects and concluded that gender differences in the language used in the written proposals lead to gender differences in success rates that are unrelated to the quality of the project or of the candidate. In our context, this finding raises the question of how to weight the three dimensions that are scored in the remote evaluation. Specifically, we ask what would be the consequences for gender inequality of altering the weights assigned to credentials, measured by candidates' academic transcript and professional experience, quality of the proposal, and letters of reference.

To address this question we provide a decomposition of the separate effects of the three factors scored during the remote evaluation. First of all, we isolate the effect of the *Transcripts & CV* component by comparing the award allocation resulting from using the aggregate score (simulation 3) to the allocation based on the other two components, namely, Proposal and Letters of reference (simulation 6). The results are presented in column 2 of Table 9. When considering all fields together, we find that *Transcripts & CV* lowers the female share among winners by 3.2 percentage points. We also find negative effects for all fields separately, except for a null effect in Social Sciences. Hence, reviewers' scores of candidates' *Transcripts & CV* increase gender inequality uniformly across all fields. This finding is not surprising because it mirrors gender differences on GPA. In the actual selection process, this score is weighted more heavily than the others, thus exacerbating gender inequality in outcomes.

Next, we turn to the role of the *Proposal* score. To isolate this factor we compare the award allocations in the scenario where both the Proposal and Letters of reference are used (simulation 6) to the scenario where awards are based solely on the Letters (simulation 5). As shown in column 3, the quality of the proposal appears to have a small positive effect on the female share among winners in general (by 1.1 percentage points), but the effect varies widely across fields. The Proposal score has a small negative effect in STEM and Social Sciences, but larger positive effects in Health & Life Sciences and in Arts & Humanities. Hence, unlike Kolev et al. (2019), we do not find evidence that the score measuring the quality of the proposal systematically penalizes female candidates.

Last, we examine the role of the *Letters*. The results in column 4 show that the score of the letters of reference has a positive effect on the female share among winners, increasing it by 2.8 percentage points. Once again, the sign of the effects masks countervailing effects across fields. In STEM, the letters play a very large role, being responsible for a 13.7 percentage-point increase in the female share among winners. In contrast, the Letters have negative effects in the female-dominated fields. The magnitudes are small in Health & Life Sciences and in Arts & Humanities, but much larger in Social Sciences. In the latter the scores of the letters account for a 7.8 percentage-point reduction in the female share among winners. Thus, the letters of reference have a large influence on the degree of gender inequality, favoring women in STEM and penalizing them in the female-dominated fields. Thus the *Letters* score is largely responsible for the *gender balancing* pattern across fields emerging in the remote evaluation, suggesting that professors (letter writers) also engage in gender balancing. This is perhaps not surprising, given that reviewers are typically chosen by the LCF from the population of university professors.

In conclusion, we find that the score for *Transcripts & CV* systematically increases gender inequality, particularly in female-dominated fields. While we do not find a systematic effect of the *Proposal* score on gender inequality, we do find that the *Letters* score is responsible for generating the pattern of *gender balancing*.

8 The Participation Decision

Above we documented that the GPA distribution of male applicants to the LCF program is shifted to the right relative to that of females. As a result, male candidates have, on average, higher GPA than female candidates. This observation is in stark contrast to the widespread finding of higher grades among female college graduates, relative to men. This striking reversal in the relative ranking of the GPA distributions of men and women suggests that there exist large gender differences in the decision to submit an application to the LCF program. To investigate this question we use our matched LCF-UB data containing all *University of Barcelona* graduates over the period 2009-2018. The merged dataset contains 432 UB graduates (2009-2018) that submitted (complete) applications to the LCF fellowship program (2014-2018).

To investigate the determinants of participation in the LCF program we estimate a linear probability model where the dependent variable ($Apply_i$) is an indicator for whether individual i participates in the program:

$$Apply_i = \alpha + \beta Fem_i + X_i'\delta + \varepsilon_i, \tag{3}$$

where vector X_i contains GPA, age, parental education, and fixed-effects for year of graduation and field of study.

Table 10 presents the main results. Column 1 estimates the model on the whole sample. As shown at the bottom of the table, the participation rate in the LCF program is low, at 0.76%. The point estimates show a positive and significant effect of GPA, indicating that interest in graduate studies is larger among students with better grades. The estimates also show a large positive effect of parental education. When both parents have college education, the probability of participating in the LCF program is 39 percentage-points larger. The coefficient for the female variable is negative, but small and imprecisely estimated.

Next, we restrict the analysis to a more relevant population, consisting of graduates with high GPA. Accordingly, column 2 restricts the analysis to the subsample of students with GPA above the 75th percentile in their respective major. The participation rate now increases by 60% reaching 1.21%. Furthermore, the coefficient for the female dummy is now much larger and significantly different from zero. More specifically, the estimated coefficient implies that female graduates are 0.4 percentage points less likely to participate in the LCF program than male candidates with the same GPA, age, socio-economic status, and in the same field of study. The coefficient on GPA is also substantially larger than in column 1. Each one-point increase in GPA (on a 0-10 scale) is associated to a 1.2 percentage-point increase in the participation rate. Column 3 further restricts the sample to graduates with GPA above the 90th percentile. The participation rate increases to 1.70% and so does the gender gap in participation, which now rises to 1.05 percentage points.

Summing up, our analysis shows that there exists a large gender gap in participation among high-GPA students. As far as we know, this is a novel finding and it compounds the gender success gap discussed earlier in the paper. To see this, it is helpful to consider a hypothetical graduating class (from the University of Barcelona) and focus on the students with GPA above the 75th percentile. How many of these men and women should be expected to obtain an LCF fellowship to fund graduate studies abroad?

The first consideration is that women currently account for 66.8% of the group under consideration.[19] Secondly, the overall probability of success for each gender can be decomposed as follows:

$$Prob(Success) = Prob(Participate) \times Prob(Success \mid Participate). \quad (4)$$

Our data allow us to estimate the terms on the right-hand side, which can be combined to compute the unconditional success probability by gender. More specifically, we compute mean participation rates by gender and field of study for UB graduating students with GPA above the 75th percentile.

As shown in Table 11, when pooling all fields, the participation rates for males and females are 1.42% and 1.12%, respectively. Thus, the female participation rate is 21% lower than the corresponding rate for males. Similarly, we estimate the mean success rate (conditional on participation) by gender, obtaining 10.92% and 7.10% for male and female candidates, respectively. Thus, the conditional success rate for women is 35% lower than for men. Column 3 computes the unconditional success probability, which is the product of the corresponding terms in columns 1 and 2. The resulting unconditional win rates are 0.16% for men and 0.08% for women, respectively. Thus, the combined gender gap is magnified and results in women's success rate falling to roughly half that of men.

The data also show interesting differences across fields of study. In STEM, the largest gender gap is found in terms of participation, with women's participation rate being 53% lower than men's.[20] In contrast in Health & Life Sciences the largest gender gap is found in the conditional success rates. Nonetheless, the *combined* gender gap is similar in both fields: women's unconditional success probability is around 1/3 of men's. In comparison, the combined gender gaps are much smaller in Social Sciences and in Arts & Humanities. In these fields the combined success rate for women is *only* 1/3 lower for women than for men.

It is interesting to simulate the effects of policies aimed at eliminating gender gaps along the participation and success margins. In the pooled data, the largest equalizing effect is obtained by closing the conditional success gap between the two genders. However, the most effective interventions differ across fields. Namely, our analysis suggests

[19] In the UB data, 66.4% of the graduates are women and this share increases to 66.8% when we restrict to students with GPA above the 75th percentile.

[20] It is important to keep in mind that the number of observations used to estimate the participation rates by gender (among high-GPA students) is relatively small in STEM, with only 676 observations. The reason is that the fraction of STEM students in overall enrollment at the UB is only around 6%. This university offers Math, Physics and other science majors but lacks engineering majors.

that closing the participation gap would have the largest effect in STEM while closing the conditional success gap would be the most effective policy in Health & Life Sciences.

9 Conclusions

In this paper we scrutinize the selection process of a highly competitive fellowship program in Spain. The structure of the process is widespread. First, all applications are reviewed remotely by experts who produce quantitative scores along a few relevant dimensions. These scores are used to rank applications and the top-ranked candidates are then invited for an in-person panel interview. Our analysis shows that the process was carefully designed and is meticulously implemented. Nonetheless, it exhibits very large gender gaps in success rates.

Our analysis has identified two factors that contribute to explain the gender inequality in outcomes. First of all, high-GPA women exhibit very low participation rates in the fellowship program. This may be due to a scarcity of female role models, low career aspirations, or high family responsibilities that reduce the interest of female graduates in pursuing graduate studies abroad. Our analysis shows that the gender gap in participation is so large that it entails a switch in the ranking of the GPA distributions of male and female graduates. As a result, there is substantial under-representation of high-achieving female candidates among the set of applicants to the fellowship program. Because the selection process is eminently meritocratic, the gender differences in GPA entail large inequality in outcomes. Future work should aim at investigating the reasons behind the low participation of high-achieving females in the fellowship program.

Secondly, we have found evidence that reviewers (as well as letter writers) engage in *gender balancing*, favoring the minority gender in each academic field. However, because women are the majority in every field except for STEM, this aspect of reviewers' behavior penalizes them. Our results underscore the findings in Rivera and Tilcsik (2019) and Kolev et al. (2019) by showing that quantitative scoring of applications is not gender neutral, but the effects of reviewers' choices are heterogeneous across fields of study.

Our analysis also suggests that reviewers should be made aware of the consequences of trying to balance the number of male and female winners in their field. Favoring the minority gender within a field may serve an important social purpose, such as creating role models that help shape the aspirations of future cohorts of graduates. But this behavior entails a trade-off as it tends to penalize female candidates in terms of their success within the program, which compounds the gender gap in participation rates.

References

Alesina, Alberto, Yann Algan, Pierre Cahuc, and Paola Giuliano, "Family values and the regulation of labor," *Journal of the European Economic Association*, 2015, *13* (4), 599–630.

Azmat, Ghazala and Barbara Petrongolo, "Gender and the labor market: What have we learned from field and lab experiments?," *Labour Economics*, 2014, *30* (C), 32–40.

_ and Rosa Ferrer, "Gender Gaps in Performance: Evidence from Young Lawyers," *Journal of Political Economy*, 2017, *125* (5), 1306–1355.

_ , Caterina Calsamiglia, and Nagore Iriberri, "Gender Differences In Response To Big Stakes," *Journal of the European Economic Association*, December 2016, *14* (6), 1372–1400.

_ , Maia Guell, and Alan Manning, "Gender Gaps in Unemployment Rates in OECD Countries," *Journal of Labor Economics*, January 2006, *24* (1), 1–38.

Bagues, Manuel F and Berta Esteve-Volart, "Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment," *The Review of Economic Studies*, 2010, *77* (4), 1301–1328.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva, "Does the Gender Composition of Scientific Committees Matter?," *American Economic Review*, April 2017, *107* (4), 1207–1238.

Bertrand, Marianne, "The Glass Ceiling," *Economica*, 2018.

_ and Esther Duflo, "Field experiments on discrimination," in "Handbook of economic field experiments," Vol. 1, Elsevier, 2017, pp. 309–393.

_ , Claudia Goldin, and Lawrence F Katz, "Dynamics of the gender gap for young professionals in the financial and corporate sectors," *American Economic Journal: Applied Economics*, 2010, *2* (3), 228–55.

_ , _ , and Lawrence F. Katz, "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors," *American Economic Journal: Applied Economics*, July 2010, *2* (3), 228–255.

_ , Emir Kamenica, and Jessica Pan, "Gender identity and relative income within households," *The Quarterly Journal of Economics*, 2015, *130* (2), 571–614.

_ , Sandra E Black, Sissel Jensen, and Adriana Lleras-Muney, "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway," *Review of Economic Studies*, 2019, *86* (1), 191–239.

Boustan, Leah and Andrew Langan, "Variation in Women's Success across PhD Programs in Economics," *Journal of Economic Perspectives*, Winter 2019, *33* (1), 23–42.

Broder, Ivy E, "Review of NSF economics proposals: Gender and institutional patterns," *The American Economic Review*, 1993, *83* (4), 964–970.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri, "Are Referees and Editors in Economics Gender Neutral?," Mimeo December 2019.

Carpio, Lucia Del and Maria Guadalupe, "More Women in Tech? Evidence from a field experiment addressing social identity," Working Papers, INSEAD 2019.

Childs, Jason, "Gender differences in lying," *Economics Letters*, 2012, *114* (2), 147–149.

Croson, Rachel and Uri Gneezy, "Gender Differences in Preferences," *Journal of Economic Literature*, June 2009, *47* (2), 448–74.

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti, "Mothers and sons: Preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 2004, *119* (4), 1249–1299.

Garcia-Montalvo, Jose, *Impacto de las becas 'La Caixa' de posgrado en el extranjero*, Obra Social 'la Caixa', 2014.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini, "Performance in competitive environments: Gender differences," *The Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

Hospido, Laura, Luc Laeven, and Ana Lamo, "The gender promotion gap: evidence from central banking," 2019.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," NBER Working Papers 18693, National Bureau of Economic Research, Inc January 2013.

Iriberri, Nagore and Pedro Rey-Biel, "Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision," *Journal of Economic Behavior & Organization*, 2017, *135* (C), 99–111.

_ and _ , "Brave Boys and Play-it-Safe Girls: Gender Differences in Willingness to Guess in a Large Scale Natural Field Experiment," CEPR Discussion Papers 13541, C.E.P.R. Discussion Papers February 2019.

_ and _ , "Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-Stage Competition in Mathematics," *The Economic Journal*, 2019.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and josef zweimueller, "Child Penalties Across Countries: Evidence and Explanations," CEPR Discussion Papers 13474, C.E.P.R. Discussion Papers January 2019.

Kolev, Julian, Yuly Fuentes-Medel, and Fiona Murray, "Is Blinded Review Enough? How Gendered Outcomes Arise Even Under Anonymous Evaluation," NBER Working Papers 25759, National Bureau of Economic Research, Inc April 2019.

Montolio, Daniel and Pere A. Taberner, "Gender differences under test pressure and their impact on academic performance: a quasi-experimental design," Working Papers 2018/21, Institut d'Economia de Barcelona (IEB) 2018.

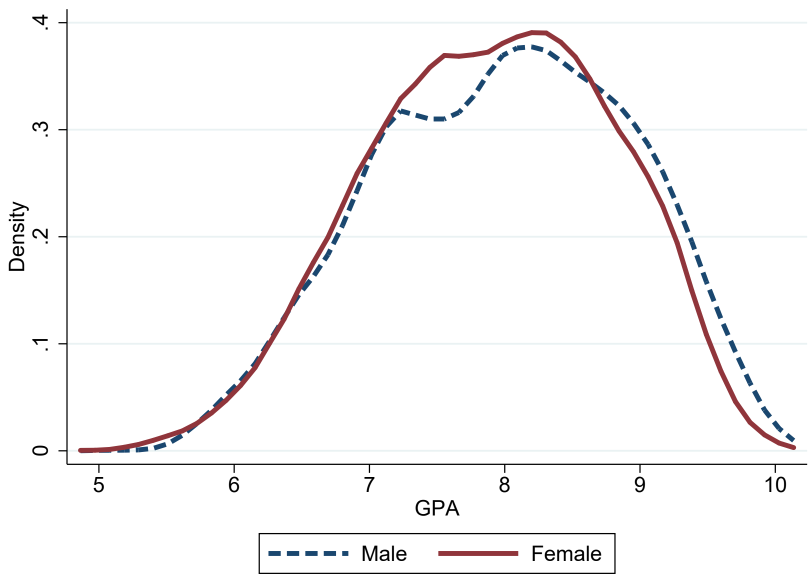
Niederle, Muriel and Lise Vesterlund, "Do women shy away from competition? Do men compete too much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

Rivera, Lauren A. and Andras Tilcsik, "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation," *American Sociological Review*, 2019, *84* (2), 248–274.

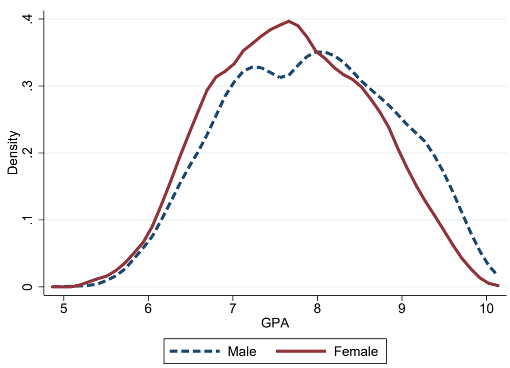
Rouse, Cecilia and Claudia Goldin, "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians," *American Economic Review*, September 2000, *90* (4), 715–741.

Todd, Sorensen, Sarnikar Surpriya, and Oaxaca Ronald L., "Do You Receive a Lighter Prison Sentence Because You Are a Woman or a White? An Economic Analysis of the Federal Criminal Sentencing Guidelines," *The B.E. Journal of Economic Analysis & Policy*, October 2013, *14* (1), 1–54.

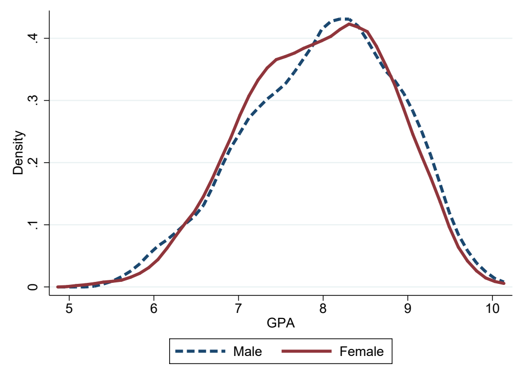
Figure 1: GPA application



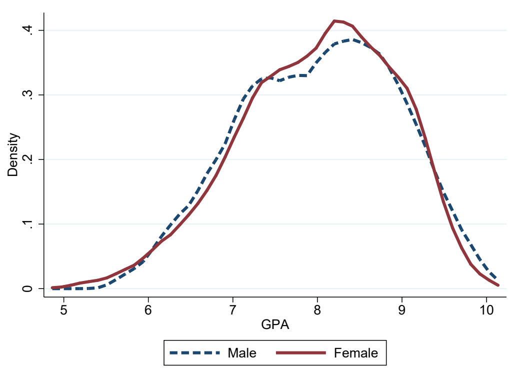
Notes: GPA in LCF application.



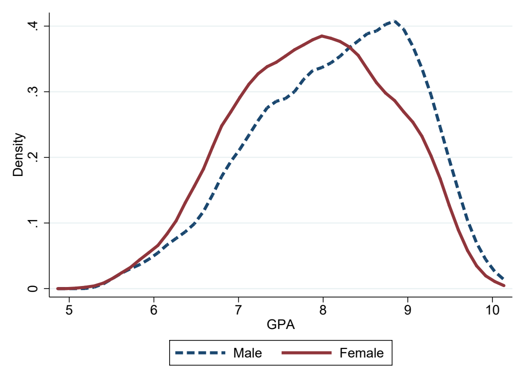
((a)) STEM



((b)) Social Sciences



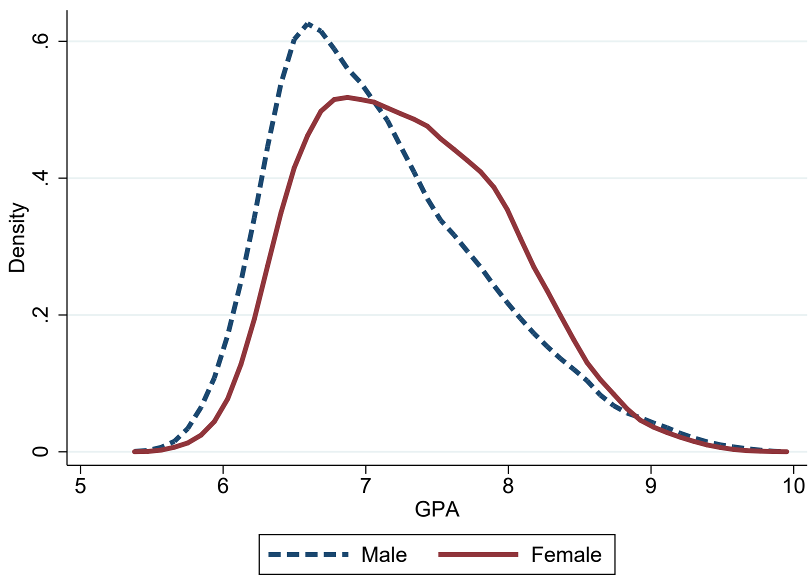
((c)) Arts & Humanities



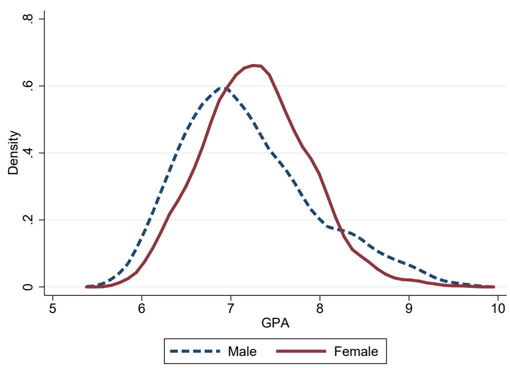
((d)) Health & Life Sciences

Figure 2: GPA distributions by gender and academic field

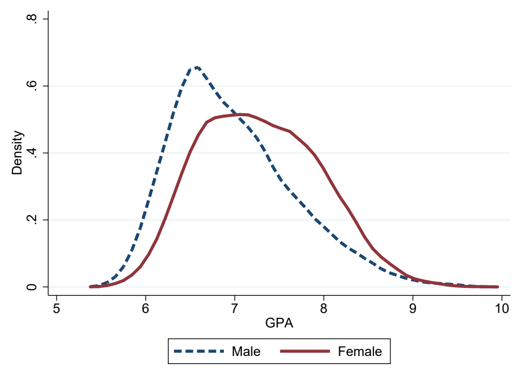
Figure 3: GPA Graduates UB



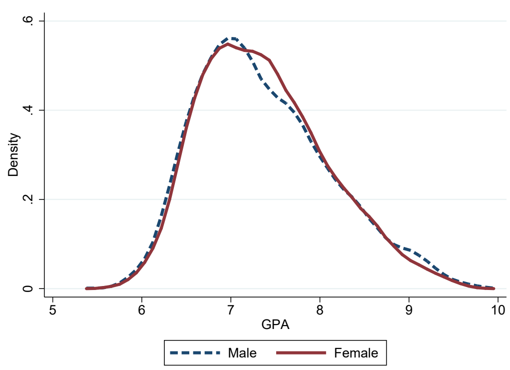
Notes: GPA in LCF application.



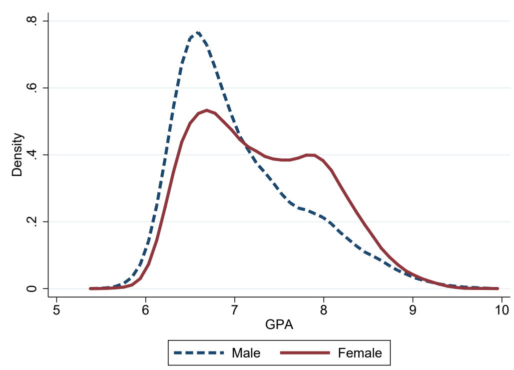
((a)) STEM



((b)) Social Sciences



((c)) Arts & Humanities



((d)) Health & Life Sciences

Figure 4: GPA distributions by gender and academic field. Graduates UB

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Success	8,099	8.828	28.372	0	100
Pass1	8,099	19.002	39.234	0	100
Female	8,099	.548	.498	0	1
Age	8,099	27.495	3.82	20	54
Age $>$ 29	8,099	.304	.46	0	1
GPA10	8,099	7.939	.913	5	10
Prog. EUR	8,099	.457	.498	0	1
Prog. AMA	8,099	.238	.426	0	1
Prog. ESP	8,099	.305	.46	0	1
STEM	8,099	.295	.456	0	1
Health & Life Sc.	8,099	.224	.417	0	1
Arts & Humanities	8,099	.222	.416	0	1
Social Sciences	8,099	.259	.438	0	1
Transcripts & CV	7,936	6.429	.923	2.9	8
Proposal	7,936	6.349	.984	2	8
Letters	8,098	6.4	.918	2.5	8

Notes: LCF applicants data. Success is an indicator for successfully completing the selection process and being offered an award. Pass1 is an indicator for advancing to the panel interview. GPA10 is the candidate's GPA on a 0-10 scale.

Table 2: Success rates

Success	(1)	(2)	(3)	(4)	(5)
Unadjusted for Grades					
Female	-3.82***	-3.84***	-3.63***	-3.45***	-3.31***
	[0.64]	[0.65]	[0.64]	[0.67]	[0.68]
Age > 29		-2.59***	-2.79***	-2.31***	-1.85**
		[0.71]	[0.71]	[0.71]	[0.72]
Grade-adjusted					
Female	-2.64***	-2.63***	-2.36***	-1.78***	-1.50**
	[0.61]	[0.61]	[0.60]	[0.63]	[0.63]
Age > 29		1.28*	1.21*	2.07***	2.54***
		[0.69]	[0.69]	[0.69]	[0.71]
GPA $50 - 75p$	4.42***	4.74***	4.88***	5.41***	5.40***
	[0.65]	[0.67]	[0.66]	[0.67]	[0.68]
GPA $75 - 90p$	14.01***	14.36***	14.79***	15.36***	15.70***
	[1.12]	[1.12]	[1.11]	[1.12]	[1.12]
GPA $> 90p$	29.26***	29.71***	30.33***	30.64***	31.43***
	[1.66]	[1.67]	[1.66]	[1.65]	[1.65]
R-squared	0.11	0.11	0.12	0.13	0.15
Observations	8,099	8,099	8,099	8,099	8,099
Mean Dep. Var.	8.83	8.83	8.83	8.83	8.83
FE year	No	Yes	Yes	Yes	Yes
FE program	No	No	Yes	Yes	Yes
FE field	No	No	No	Yes	Yes
FE origin univ.	No	No	No	No	Yes

Notes: The dependent variable has been multiplied by 100 to re-scale coefficients. All models include a dummy for age of the applicant above 29 years. Heteroskedasticity-robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Success rates (2). Adjusted by grades and scores remote evaluation

Success	(1)	(2)	(3)	(4)	(5)
Female	-1.55***	-1.53***	-1.31**	-1.22**	-1.09*
	[0.59]	[0.59]	[0.58]	[0.61]	[0.61]
GPA $50-75p$	-1.55**	-1.37**	-1.16*	-0.99	-0.71
	[0.68]	[0.69]	[0.69]	[0.69]	[0.71]
GPA $75-90p$	4.12***	4.28***	4.82***	5.00***	5.68***
	[1.11]	[1.11]	[1.09]	[1.10]	[1.12]
GPA $> 90p$	16.72***	16.91***	17.67***	17.79***	18.82***
	[1.72]	[1.72]	[1.69]	[1.70]	[1.71]
Transcripts & CV	5.96***	6.05***	5.70***	5.76***	5.43***
	[0.48]	[0.48]	[0.47]	[0.47]	[0.48]
Proposal	2.85***	2.79***	2.83***	2.72***	2.70***
	[0.35]	[0.35]	[0.35]	[0.36]	[0.36]
Letters	1.82***	1.88***	2.32***	2.23***	2.23***
	[0.37]	[0.37]	[0.37]	[0.38]	[0.38]
Observations	7,936	7,936	7,936	7,936	7,936
R-squared	0.18	0.18	0.20	0.20	0.22
Mean Dep. Var.	8.83	8.83	8.83	8.83	8.83
FE year	No	Yes	Yes	Yes	Yes
FE program	No	No	Yes	Yes	Yes
FE field	No	No	No	Yes	Yes
FE origin univ.	No	No	No	No	Yes

Notes: The dependent variable has been multiplied by 100 to re-scale coefficients. All models include a dummy for age of the applicant above 29 years. During the pre-selection each application is scored by two reviewers along 3 dimensions: (i) Transcripts and CV, (ii) Quality of the proposal, and (iii) Letters of reference. Each dimension is scored from 1 to 8. The specifications here include the average of the scores of the two reviewers for each of the dimensions. Heteroskedasticity-robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Reviews - Remote evaluation

Dep. Var.	1 Transcripts&CV	2 Proposal	3 Letters	4 Transcripts&CV	5 Proposal	6 Letters
Unadjusted						
Female	-0.16*** [0.01]	-0.08*** [0.02]	-0.08*** [0.01]	-0.17*** [0.02]	-0.05* [0.02]	-0.06*** [0.02]
$Fem \times RevFem$				0.01 [0.03]	-0.07** [0.03]	-0.03 [0.03]
Grade-adjusted						
Female	-0.06*** [0.01]	-0.04** [0.02]	-0.04*** [0.01]	-0.07*** [0.02]	-0.01 [0.02]	-0.02 [0.02]
$Fem \times RevFem$				0.01 [0.02]	-0.06** [0.03]	-0.04 [0.03]
GPA $50-75p$	0.77*** [0.01]	0.34*** [0.02]	0.35*** [0.02]	0.77*** [0.01]	0.34*** [0.02]	0.35*** [0.02]
GPA $75-90p$	1.21*** [0.02]	0.59*** [0.02]	0.57*** [0.02]	1.21*** [0.02]	0.59*** [0.02]	0.57*** [0.02]
$GPA > 90p$	1.56*** [0.02]	0.77*** [0.03]	0.81*** [0.02]	1.56*** [0.02]	0.77*** [0.03]	0.82*** [0.02]
R-squared	0.36	0.10	0.11	0.36	0.10	0.11
Obs.	18,019	18,019	18,439	17,918	17,918	18,338
Mean dep. var.	6.37	6.42	6.46	6.37	6.42	6.46
FE origin univ.	Yes	Yes	Yes	Yes	Yes	Yes
FE acad. field	Yes	Yes	Yes	Yes	Yes	Yes
FE program	Yes	Yes	Yes	Yes	Yes	Yes
FE reviewer	Yes	Yes	Yes	Yes	Yes	Yes
FE year	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All specifications control for age by including a dummy variable for age above 29 years old, the 75th percentile among applicants to the program. All programs (AMA, EUR, ESP) pooled. Each applications was reviewed remotely by two reviewers. In total 323 reviewers. Heteroskedasticity-robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Success by Academic Field

Success	1 STEM	2 Health & Life S.	3 Humanities	4 Social Sciences
Unadjusted				
Female	-1.37	-7.75***	-1.92	-2.03
	[1.28]	[1.61]	[1.44]	[1.25]
Grade-adjusted				
Female	1.44	-5.15***	-1.25	-1.20
	[1.19]	[1.44]	[1.40]	[1.18]
GPA 50-75p	7.27***	4.55***	3.44**	5.51***
	[1.46]	[1.40]	[1.42]	[1.24]
GPA 75-90p	21.80***	13.25***	10.27***	16.43***
	[2.53]	[2.26]	[1.99]	[2.23]
GPA $> 90p$	36.62***	37.16***	23.11***	27.46***
	[3.11]	[3.41]	[3.25]	[3.57]
Mean Dep. Var.	10.4	9.16	8.12	7.34
Observations	2,355	1,792	1,754	2,035
FE year	Yes	Yes	Yes	Yes
FE program	Yes	Yes	Yes	Yes
FE origin univ.	Yes	Yes	Yes	Yes

Notes: The dependent variable has been multiplied by 100 to re-scale coefficients. All models include a dummy for applicants older than 29. Heteroskedasticity-robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Reviewer Scores by Academic Field. Remote evaluation

Dep. Var.	Transcripts&CV	Proposal	Letters
All Fields, N=	18,019	18,019	18,439
Mean dep. var	6.46	6.37	6.42
Female	-0.06***	-0.04**	-0.04***
	[0.01]	[0.02]	[0.01]
STEM, N=	5,280	5,280	5,386
Mean dep. var	6.53	6.52	6.50
Female	-0.01	0.09***	0.13***
	[0.02]	[0.03]	[0.03]
Life & Health, N=	3,993	3,993	4,051
Mean dep. var	6.53	6.58	6.67
Female	-0.05*	-0.04	-0.06**
	[0.02]	[0.03]	[0.03]
Arts&Hum., N=	4,173	4,173	4,273
Mean dep. var	6.48	6.17	6.27
Female	-0.08***	-0.05	-0.10***
	[0.03]	[0.04]	[0.03]
Social Sci., N=	4,573	4,573	4,729
Mean dep. var	6.40	6.23	6.28
Female	-0.04	-0.08**	-0.07**
	[0.02]	[0.03]	[0.03]

Notes: All models include reviewer fixed-effects, indicators for GPA in percentiles 50-75, 75-90 or above 90th percentile, and fixed-effects by year, by program and by university of origin. The mean success rates in stage 1 are 20.31 (All Fields), 22.94 (STEM), 20.53 (Life & Health), 18.94 (Arts & Humanities), 18.53 (Social Sciences). Standard errors are clustered by reviewer. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7: Simulated award allocations

Female Share winners	Actual	Sim1	Sim2	Sim3	Sim4	Sim5	Sim6
Criteria	All	GPA	T&CV	Score1	Proposal	Letters	Prop&Letters
All Fields, N=8,099	715	715	715	715	715	715	715
Actual award (pct)	100	37.3	53.6	57.3	45.6	42.8	46.2
Score1	7.4	7.2	7.5	7.6	7.5	7.4	7.5
FSH Candidates	54.8	54.8	54.8	54.8	54.8	54.8	54.8
FSH Winners	44.1	43.2	44.1	43.9	45.45	46.0	47.1
FSH Winners - Candidates	-10.7	-11.6	-10.7	-10.9	-9.3	-8.8	-7.7
STEM, N=2,393	249	249	249	249	249	249	249
FSH Candidates	34.4	34.4	34.4	34.4	34.4	34.4	34.4
FSH Winners	27.3	17.3	25.3	28.1	26.91	30.9	29.3
FSH Winners - Candidates	-7.0	-17.1	-9.0	-6.2	-7.4	-3.4	-5.0
Health & Life Sc., N=1,812	166	166	166	166	166	166	166
FSH Candidates	67.6	67.6	67.6	67.6	67.6	67.6	67.6
FSH Winners	48.8	56.0	53.0	52.4	55.42	55.4	57.8
FSH Winners - Candidates	-18.8	-11.6	-14.6	-15.2	-12.2	-12.2	-9.8
Social Sciences, N=2,097	146	146	146	146	146	146	146
FSH Candidates	59.6	59.6	59.6	59.6	59.6	59.6	59.6
FSH Winners	51.3	56.5	51.9	47.4	44.81	48.7	47.4
FSH Winners - Candidates	-8.3	-3.1	-7.6	-12.2	-14.8	-10.9	-12.2
Arts & Hum., N=1,797	154	154	154	154	154	154	154
FSH Candidates	63.5	63.5	63.5	63.5	63.5	63.5	63.5
FSH Winners	59.6	58.9	57.5	57.5	66.44	58.2	65.1
FSH Winners - Candidates	-3.9	-4.6	-6.0	-6.0	2.9	-5.3	1.6

Notes: Column 1 reports data based on the actual allocation of awards (Data). The following columns report figures based on simulated allocations of awards based on the criteria specified in the Table. Simulation 3 is based on a score that aggregates all pre-selection scores according to the following formula: $Score1 = 0.5 * Transcripts\&CV + 0.3 * Proposal + 0.2 * Letters$. Simulation 6 aggregates the pre-selection scores for the quality of the proposal and the letters of reference with weights 0.6 and 0.4, respectively.

Table 8: Decomposition remote evaluation vs. panel interview

Δ Fem. Sh. Win Field (% Fem)	Full Process Actual - Sim1	Interview Actual - Sim3	Remote evaluation Sim3 - Sim1
All Fields (55%)	0.8	0.1	0.7
STEM (34%)	10.0	-0.8	10.8
Health & Life (68%)	-7.2	-3.6	-3.6
Social Sciences (60%)	-5.2	3.9	-9.1
Arts & Humanities (63%)	0.7	2.1	-1.4
SumAbs across fields	23.1	10.4	24.9

Notes: Computations based on the simulations reported in Table 7. Each panel reports the change in the female share among winners (Δ FShWin), defined as the ratio of female winners to female candidates. The bottom panel reports the sum of the absolute value of the effect in each column across the 4 fields of study (excluding the effect on the pooled data).

Table 9: Decomposition scores remote evaluation

Δ Fem. Sh. Win Field (% Fem)	Remote Eval. Sim3 - Sim1	Trans & CV Sim3-Sim6	Proposal Sim6-Sim5	Letters Sim5-Sim1
All Fields (55%)	0.7	-3.2	1.1	2.8
STEM (34%)	10.8	-1.2	-1.6	13.7
Health & Life (68%)	-3.6	-5.4	2.4	-0.6
Social Sciences(60%)	-9.1	0.0	-1.3	-7.8
Arts & Humanities (63%)	-1.4	-7.5	6.8	-0.7
SumAbs FShW across fields	24.9	14.2	12.2	22.7

Notes: Computations based on the simulations reported in Table 7. Each panel reports the change in the female share among winners (Δ FShWin), defined as the ratio of female winners to female candidates. The bottom panel reports the sum of the absolute value of the effect in each column across the 4 fields of study (excluding the effect on the pooled data).

Table 10: The participation decision in LCF program. UB graduates

	(1)	(2)	(3)
Dep. Var.	Apply	Apply	Apply
GPA	All	$GPA > 75p$	$GPA > 90p$
Female	-0.09	-0.42**	-1.05***
	[0.08]	[0.20]	[0.37]
GPA	0.50***	1.19***	1.69***
	[0.05]	[0.16]	[0.34]
Age	-0.02***	-0.06***	-0.08***
	[0.01]	[0.01]	[0.03]
One parent College	0.03	0.04	-0.10
	[0.09]	[0.24]	[0.44]
Both parents College	0.39***	0.48**	0.61
	[0.10]	[0.23]	[0.43]
Observations	56,946	14,223	5,696
R-squared	0.00	0.01	0.01
Mean dep. Var.	0.76%	1.21%	1.70%
FE Year	yes	yes	yes
FE field	yes	yes	yes

Notes: The sample includes all the *University of Barcelona* graduates during academic years 2009-2018. The dependent variable is an indicator taking the value of 100 when the student applied to the LCF fellowship program (during the period 2014-2018) and zero otherwise.

Table 11: Simulation participation and success by field

Probability	(1) Baseline Participate (%)	(2) Baseline Win | *Participate*(%)	(3) Baseline Win (%)	(4) No gap partic. Win (%)	(5) No gap success Win (%)
All Fields					
Males	1.42	10.92	0.16	0.13	0.13
Females	1.12	7.10	0.08	0.09	0.10
All	1.22	8.83	0.11	0.11	0.11
Fem/Male	0.79	0.65	0.51	0.65	0.79
STEM					
Males	1.68	11.52	0.19	0.14	0.18
Females	0.79	8.27	0.07	0.10	0.08
All	1.18	10.41	0.12	0.12	0.12
Fem/Male	0.47	0.72	0.34	0.72	0.47
Health & Life Sc.					
Males	1.81	14.48	0.26	0.24	0.17
Females	1.56	6.61	0.10	0.11	0.14
All	1.65	9.16	0.15	0.15	0.15
Fem/Male	0.86	0.46	0.39	0.46	0.86
Social Sc.					
Males	0.90	8.99	0.08	0.07	0.07
Females	0.72	7.62	0.06	0.06	0.06
All	0.77	8.12	0.06	0.06	0.06
Fem/Male	0.80	0.85	0.68	0.85	0.80
Arts & Humanities					
Males	1.83	8.84	0.16	0.15	0.13
Females	1.67	6.33	0.11	0.11	0.12
All	1.73	7.34	0.13	0.13	0.13
Fem/Male	0.92	0.72	0.66	0.72	0.92

Notes: The simulation is based on the population of college graduates with high grades, defined as GPA above the 75th percentile in their MAJOR AND UNIVERSITY. CHECK. The mean participation rates are based on the UB-LCF dataset. The mean conditional success rates are obtained from the LCF dataset. These means only condition by gender and 75th percentile in the GPA distribution. Columns 1-3 (baseline) are based on data. Column 4 is a simulation that assumes the same participation for males and females, given by the mean participation rate in the corresponding field of study (pooling both genders). Column 4 is a simulation that assumes the same success conditional in participation in the LCF program for both genders, given by the mean success rate in the corresponding field.

Appendix

Table 12: Simulations. Success rates

Success rate Criteria	Actual All	Sim1 GPA	Sim2 T&CV	Sim3 Score1	Sim4 Proposal	Sim5 Letters	Sim6 Prop&Letters
All Fields, N=8,099	715	715	715	715	715	715	715
Actual award (pct)	100	37.34	53.57	57.34	45.59	42.80	46.15
Score1	7.43	7.20	7.52	7.63	7.50	7.45	7.54
Males	10.92	11.09	10.92	10.95	10.65	10.54	10.32
Females	7.10	6.96	7.10	7.08	7.32	7.41	7.60
Fem/Males	0.65	0.63	0.65	0.65	0.69	0.70	0.74
STEM, N=2,393	249	249	249	249	249	249	249
Males	11.52	13.11	11.84	11.39	11.58	10.95	11.20
Females	8.27	5.23	7.66	8.52	8.15	9.37	8.88
Fem/Males	0.72	0.40	0.65	0.75	0.70	0.86	0.79
Health & Life Sc., N=1,812	166	166	166	166	166	166	166
Males	14.48	12.44	13.29	13.46	12.61	12.61	11.93
Females	6.61	7.59	7.18	7.10	7.51	7.51	7.84
Fem/Males	0.46	0.61	0.54	0.53	0.60	0.60	0.66
Social Sc., N=2,097	146	146	146	146	146	146	146
Males	8.84	7.90	8.73	9.55	10.02	9.32	9.55
Females	6.33	6.97	6.41	5.84	5.52	6.00	5.84
Fem/Males	0.72	0.88	0.73	0.61	0.55	0.64	0.61
Arts & Hum., N=1,797	154	154	154	154	154	154	154
Males	8.99	9.15	9.45	9.45	7.47	9.30	7.77
Females	7.62	7.54	7.36	7.36	8.50	7.45	8.33
Fem/Males	0.85	0.82	0.78	0.78	1.14	0.80	1.07

Notes: Column 1 reports data based on the actual allocation of awards (Data). The following columns report figures based on simulated allocations of awards based on the criteria specified in the Table. Simulation 3 is based on a score that aggregates all pre-selection scores according to the following formula: $Score1 = 0.5 * Transcripts\&CV + 0.3 * Proposal + 0.2 * Letters$. Simulation 6 aggregates the pre-selection scores for the quality of the proposal and the letters of reference with weights 0.6 and 0.4, respectively.