Promoting Female Talent in Science: Evidence from an Affirmative Action Policy

Lídia Farré (IAE-CSIC & BSE)
Judit Vall-Castelló (UB & IEB)

January 2025

Abstract

This paper evaluates an affirmative action (AA) policy that granted 0.25 out of 10 evaluation points to research groups in Catalonia (Spain) led by a female principal investigator (PI). We use rich administrative data on the universe of applications in three consecutive funding calls for research groups (2014, 2017 and 2021) to show that, after the implementation of the policy (in 2021), the share of groups led by women increased by 8 percentage points. The increase was largely driven by high-quality groups that substituted a male by a female PI. We also show that the policy eliminates the unexplained pre-reform penalties in the evaluation process against groups led by females, increasing their chances of getting funded. Finally, we show that AA is more effective (and less costly) in promoting female talent than other more conventional policies, such as imposing a minimum share of females in the review board.

Keywords: Gender gaps in science, affirmative action policies, female talent, research on research
JEL Codes: J16, I23, I28, J78

We are grateful to the AGAUR for providing access to the data and for their overall support. In particular, we appreciate the support and advice offered by Alberto Córdoba and Victòria Miquel. We also thank seminar participants at ESPE 2024 and Simposio 2024. We acknowledge financial support from the Ministry of Science and Innovation (grant PID2022-1402060B-I00) and the Severo Ochoa Program for Centers of Excellence in R and D (BSE CEX2019-000915-S) funded by MCIN/AEI/10.13039/501100011033.

Introduction

Women are underrepresented in the scientific profession, with the gender gap widening at higher career levels. The lack of women in leadership positions can restrict diversity in the content of the research being carried out, potentially overlooking issues that significantly affect half of the population. Furthermore, the gender gap in leadership positions can perpetuate a culture of bias and unequal access to mentorship, funding, and recognition, ultimately hindering efficiency in the scientific profession.

In the European Union, women comprise 48.1% of doctoral graduates but represent only 26% of Grade A academic staff, the highest level of research positions (European Commission, 2021). Comparable trends are observed in the United States and Canada, where women hold approximately 40% of all tenured or tenure-track faculty positions across disciplines, with this number significantly declining in the senior academic ranks, as only about 26% (32%) of full professors are women in the United States (Canada) (Canadian Association of University Teachers, 2020; National Center for Education Statistics, 2022).

There are several policies that could be effective in reducing these gender gaps and this paper examines the impact of one of them: an affirmative action policy to promote female talent in science. Introduced in Catalonia (Spain) in 2021, the policy modified the allocation criteria for public funding among research groups at universities and research institutions. Specifically, it awarded additional points (0.25 out of 10) to research groups led by women, as well as to those with at least 40% female researchers (an additional 0.25 out of 10). Using rich administrative data from AGAUR (Agency for Management of University and Research Grants) on the universe of applications to two consecutive pre-reform and one post-reform calls (2014, 2017 and 2021), we first examine the impact of the policy on the composition of research groups that apply to the call. Next, we identify the effect of the reform on the selection process and the promotion of female leadership.

Our results indicate that, after the implementation of the policy, the percentage of applications led by female researchers increased. A comparison between traditionally male-dominated (STEM) and female-dominated (non-STEM) fields, shows that the policy raised the share of female led groups by at least 8 percentage points. This increase is mainly driven by high-quality groups changing from a male to a female PI. We also observe a reduction in the share of male leaders in the newly created groups.

A detailed examination of the group characteristics across the entire sample period

reveals that groups led by females are different. On average, female PIs are younger and less likely to be full professors. These groups are also smaller in size, concentrate a higher share of female researchers, and accumulate less publications and competitive research projects. However, in the post-reform call, we document an increase in the size and the quality of groups led by females.

In order to examine the content of the research carried out by each type of group, we proceed with a text analysis of the almost 4300 applications. With all the information included in the text of the proposal, we apply machine learning models to construct variables that identify the research projects that include a gender perspective. Analyzing these constructed measures, we show that groups led by females are 7 percentage points more likely to include the gender perspective dimension in their research. Our results also indicate that, in the post-reform call, there is an increase in the probability that groups led by males incorporate a gender perspective.

Finally, we investigate the effect of the policy on the dynamics of the selection process, as well as on the probability of receiving funding. Once we control for observable differences in the quality and composition of groups, we show that, in the pre-reform calls, female-led research groups received 0.13 points less in the evaluation process and were 12 percentage points less likely to be selected for funding, when compared to equivalent male-led groups. With the policy, the gender gaps in the allocation of funding are completely eliminated, as the unexplained gender penalties in scores are compensated by the AA policy.

Therefore, we conclude that the AA policy equalizes the chances of receiving funding among female and male scientific research leaders. Our analysis also highlights that the AA policy is more effective (and less costly) in promoting female talent in science than more conventional policies, such as increasing the share of women in the evaluation panels (which have proven to be quite ineffective for all females, and rather costly for senior women).

Our findings are highly relevant to the pressing concern surrounding female underrepresentation in the scientific profession. The absence of women researchers becomes more pronounced at advanced career stages due to the "leaky pipeline" and "glass ceiling" phenomena. The "leaky pipeline" problem refers to the progressive attrition of women from the academic and research career paths at various stages. It has been extensively documented in the previous literature (Kahn, 1995; Ginther and Kahn, 2004; Lundberg and Stearns, 2019; Bateman and Hengel, 2023; Ayarza and Iriberri, 2024; Iaria, Schwarz, and Waldinger, 2024) and it implies that, increasing the proportion of graduate or early-

career women, does not necessarily translate into proportional representation at higher academic ranks.

The "glass ceiling" refers to structural barriers, such as discrimination, gender biases, and unequal access to opportunities, which hinder women's advancement to top decision-making and leadership positions in academia and research institutions. A recent literature has identified several dimensions in which these barriers emerge. In the publication process, women face longer review times, they are held to higher standards, and have lower acceptance rates, suggesting potential biases in editorial practices (Card et al., 2020; Hengel, 2022; Alexander et al., 2023). Similarly, research on conference acceptance, such as Chari and Goldsmith-Pinkham (2017) and Hospido and Sanz (2021), finds that papers of female authors are less likely to be accepted for presentation, even when controlling for quality measures. In the recognition of collaborative work, Sarsons et al. (2021) show that women receive less credit for coauthored publications vis-a-vis their male counterparts. Studies on job applications in the economics job market Baltrunaite, Casarico, and Rizzica (2024) and Eberhardt, Facchini, and Rueda (2023), find that recommendation letters for women are more likely to emphasize interpersonal traits over research achievements, potentially disadvantaging them in hiring decisions.

Broader professional dynamics are also subject to these differences between men and women. **Wu (2020)**, Dupas et al. (2021), Handlan and Sheng (2023), and Seré (2023) investigate the professional climate during seminars, revealing that women are more likely to face interruptions and hostile questioning. In citation patterns, Koffi (2021) identifies gender disparities in the frequency and visibility of women's work being cited, which affects their professional recognition and career advancement. Similarly, Card et al. (2022) and Card et al. (2023) examine gender imbalances in peer recognition (like access to prestigious societies, such as the Econometric Society) documenting historically negative penalties for women, turning to positive premiums in recent years. Other studies focus on visibility and representation. Venus (2024) highlights the underrepresentation of female economists on platforms like Wikipedia, while Funk, Iriberri, and Venus (2024) document the scarcity of women in editorial roles in leading economic journals.

These gaps in academia are not only present in research but also in teaching tasks. Boring (2017) and Mengel, Sauermann, and Zölitz (2019) show that female instructors receive lower teaching evaluations, with penalties increasing for younger women in charge of technical or quantitative courses. Finally, there is also evidence that women are asked more and accept more often to participate in non-promotable tasks (Babcock et al., 2017).

Our paper contributes to this growing body of literature on gender disparities in the

scientific profession by focusing on a new dimension, access to research funding. Research funds are a critical input to career advancement and represent an important element for the evaluation of research performance. Gender differences in funding success rates will create a stumbling block for the career advancement of women, acting as a reinforcing tool of the already large gender biases in academia. Lower success in funding applications for women researchers will lead to fewer publications and patents which, in turn, will reduce the competitiveness and the probability of success of future funding applications. According to recent data published by the European Commission, the funding success rate for male applicants is 3.9 percentage points higher than that for female applicants (European Commission, 2021). In an ongoing project, Fuchs-Schündeln and Holub (work in progress) examine the presence of gender gaps in the ERC Evaluation process. In **Farré et al. (2024)** we also identify that ERC projects that focus on types of cancers that have a larger incidence on females than males (female dominated cancers) receive less funding than those that focus on male-dominated cancers.

A second important contribution of our paper is the analysis of a policy that has the objective of reducing the observed gender gaps in academia and, more specifically, in leadership positions in research. The existence of these gaps has become a significant concern from both an equity and efficiency perspective, prompting the implementation of various policies aimed at reducing these disparities (see Buckles (2019) for a survey). One of the early policy initiatives to address gender imbalances in hiring and promotion decisions was the establishment of a minimum share of women on evaluation committees. While this policy has been widely adopted, it has placed a considerable burden on senior women, who are often asked to serve on numerous committees. Furthermore, the existing evidence shows that the impact of these kind of policies in reducing gender gaps remains limited M. F. Bagues and Esteve-Volart (2010), M. Bagues, Sylos-Labini, and Zinovyeva (2017), and Farré and Ortega (2023). In the publication process, efforts to increase female representation on editorial boards have significantly yielded limited success in narrowing gender gaps in academic publishing (Card et al., 2020) **(Hengel 2020)**. Other policy initiatives aimed at promoting female talent in academia include gender-neutral clock-stopping policies **(Antecol et al., 2018)**, variations in peer review processes (Tomkins, Zhang, and Heavlin, 2017), mentoring programs (Ginther et al., 2020) and raising awareness of gender disparities in the profession (Card et al., 2022; Gamage and Sevilla, 2019). However, the evidence suggests that these efforts are only bringing modest improvements in gender equity in academia, as their effects on the advancement of female talent have proven to be relatively mild.

In our paper, we focus on a relatively unexplored policy, affirmative action (AA), that has the potential to effectively reduce these gaps. In general, affirmative action

policies are designed to increase the representation of groups in areas where they have traditionally been excluded or where they face systematic barriers. In academia, these policies typically involve giving preference to qualified female candidates in hiring and promotion decisions or setting targets for female representation in leadership positions. Accordingly, AA policies have an immediate redistributive effect by directly increasing the presence of women in historically underrepresented positions. Apart from improving the chances of women, these policies can also have additional effects, such as enhancing visibility. Although in a very different context, Beaman et al. (2009) find that exposure to female leaders in India, not only reduced biases against women leadership roles, but also increased the aspirations of young girls, illustrating the potential long-run benefits of increased female representation. AA policies have been a popular tool to increase the representation of ethnic minorities and disadvantaged groups in college participation (Arcidiacono and Lovenheim, 2016; Mello, 2023). In a recent study, Bleemer (2022) evaluates the effects of Prop 209 in California, which banned AA in public university admissions. The results show a decline in the number of Black and Hispanic applicants, which affected their chances of being admitted, but they also highlight how AA policies contribute to greater socioeconomic mobility of underrepresented groups.

Even with all these potential benefits from AA polices, there are also efficiency concerns associated with these types of interventions. In their extensive survey, Holzer and Neumark (2000) review several cases of AA implementation and conclude that the empirical case against affirmative action on efficiency is weak, at best. Their study shows lower educational qualifications among women and minorities hired under affirmative action but, for most groups, they do not find evidence of weaker job performance.

To the best of our knowledge, this is the first study to investigate the impact of an affirmative action (AA) policy on the promotion of female talent in academia and, more specifically, in research funding. As our results show that the AA policy successfully reduced the unexplained gender penalties for female PIs in the funding review process, we conclude that the policy had substantial redistributive effects without damaging efficiency. Our results align with Balafoutas and Sutter (2012), whose laboratory experiment on hiring decisions found that AA increased the number of women hired without negatively affecting overall performance. This suggests that the concerns about AA leading to less-qualified hires may be overstated. Furthermore, our results suggest that AA policies may be more cost-effective in promoting female leadership among academics than other interventions, such as imposing a mandatory share of female reviewers in the evaluation process.

The paper is structured as follows. The next section provides details on the institu-

tional context and describes the data. The third section investigates the impact of the AA policy on group composition, followed by a detailed examination of its implications for the evaluation process. Finally, we conclude with some remarks.

Institutional context and data

Catalonia is home to top scientific institutions (Institute of Chemical Research, Institute of Photonic Sciences, Alba Synchrotron, Barcelona Supercomputing Center, Barcelona School of Economics, among others). In 2014, the Catalan government introduced a program to recognize and provide funding to research groups in the territory (Suport Groups de Recerca, SGR). The program grants around 50.000 euros (with a maximum of 75,000 euros) for 3 years to research groups that are positively evaluated by the program reviewers, based on quality and merits. The program has a high participation rate among researchers from all disciplines who are based in Catalonia. In 2021, the government introduced an affirmative action policy to this recognition program with the explicit aim of fostering female leadership in research. More specifically, the newly introduced rules stated that 0.25 out of 10 points were reserved for groups led by female PI's, and another 0.25 points (again, out of 10) were reserved to groups with, at least, 40% of female researchers.

Since its first implementation in 2014, the selection process is done in two steps. The first one is entirely remote, and each application is evaluated by three external experts that belong to the specific scientific area of the proposal. The three experts are chosen from a pool of researchers that the Research Agency has for each area. The scientific area of research is chosen by the PI in the application process. Each of the three external reviewers reads and grades the proposal. In the second step, one evaluation panel for each of the eight areas of research (Arts and Humanities, Biomedicine, Sciences, Medical and Health Sciences, Natural and Agrofood Sciences, Social Sciences, Engineering Technology and Architecture, Valorization and Transference) will meet to revise the results of all the projects in the first step and to issue a final evaluation proposal. These evaluation panels include external experts from different sub-fields that belong to the same area of research. The experts have a very strong international research trajectory and will bring a multidisciplinary vision to the panel. The specific role of the panel of experts is to revise the evaluations given by the remote reviewers in the first step, particularly those applications that have contradictory grades from the three external reviewers, and to issue a final evaluation proposal with a single grade for each project, integrating the evaluation grades of the three external reviewers. The panel is also in charge of setting the threshold grade that will define the granted and non-granted projects. In the two steps, grades are expressed on a scale from 1 to 10. As the literature has shown several

gender differences between STEM and non-STEM fields, we group the eight fields of research into these two categories. In both steps, the external reviewers and the experts from the evaluation panels are given some instructions to follow during the evaluation process. Apart from checking for the existence of conflicts of interest, the guidelines are also highlighting that the focus should be placed on several characteristics of the group, such as scientific productivity, the ability to gain competitive research projects, knowledge transfer and science socialization activities, as well as the scientific activities proposed for the following three years.

We have rich administrative data that includes all the information that each of the reviewers involved in the two steps have. Therefore, we are able to account for all the same observable characteristics that the reviewers take into consideration when evaluating the proposals. More specifically, our data contains several important elements on the composition and characteristics of the research groups applying for funding in the three calls (2014 with 1,652 groups; 2017 with 1,368 groups and 2021 with 1,272 groups), those groups that were granted funding, as well as detailed information on the gender of the evaluators and the grades given by each of them to each of the applications. Table 1 presents descriptive statistics for all groups and separately for STEM (i.e. engineering and architecture, science, health and medicine) and non-STEM groups (i.e. social science and humanities). The first row in the Table, which shows the share of female researchers, indicates that the presence of women in the scientific profession in Catalonia has remained relatively stable over the period under analysis. Women represented 46% of the research body in 2014, and this number increased slightly to 48% in 2021. As expected, the numbers are lower in traditionally male-dominated STEM fields, 45% in 2014 and 46% in 2021, than in non-STEM fields, 48% in 2014 and 52% in 2021.

Despite the high presence of women among researchers, they are clearly underrepresented in leadership positions. In 2014 and 2017, only 27% and 29% of the groups that applied for funding were led by females. After the policy was introduced in 2021, the share of female-led groups increased to 41%. A comparison between traditionally male-dominated STEM and female-dominated non-STEM fields allows us to gain some insights on the effect of the policy on leading female scientists. Note that in non-STEM fields, the presence of women in leadership position was relatively high before the implementation of the policy. In these fields, the share of female PIs was 40% in 2014 and 2017. This share increased by 7 percentage points, to 47% after the introduction of the policy in 2021. In contrast, the share of female PIs in STEM was much lower before the policy, 22% in 2014 and 25% in 2017, and it increased to 40% in 2021, which represents a 15 percentage points increase. Thus, these numbers suggest that the affirmative action policy had a stronger effect for female leaders in STEM fields.

Table 1 also provides descriptive statistics of the outcomes of the selection process. The fourth row in the table displays the share of groups led by females that received funding. Both in 2014 and 2017 this share was slightly below that of the female-led groups applying for funding 24% vs 27% in 2014 and 26% vs 29% in 2017. However, in 2021 these number reverted and 44% of the groups that received funding were led by women, while they only represented 41% of the total number of groups that applied for funding.

In 2017, female-led groups scored on average 0.32 points lower than male-led groups, with a slightly larger gap in non-STEM fields (-0.33) compared to STEM fields (-0.28). By 2021, before accounting for the AA points, the gap had narrowed by approximately 50%, reducing to -0.15 across all fields (-0.15 in STEM and -0.14 in non-STEM fields). This reduction may reflect changes in the quality of female-led groups applying for funding or a shift in reviewer behavior in response to the implementation of policies aimed at promoting female leadership and increasing awareness of gender disparities. The following section investigates these potential mechanisms in greater detail.

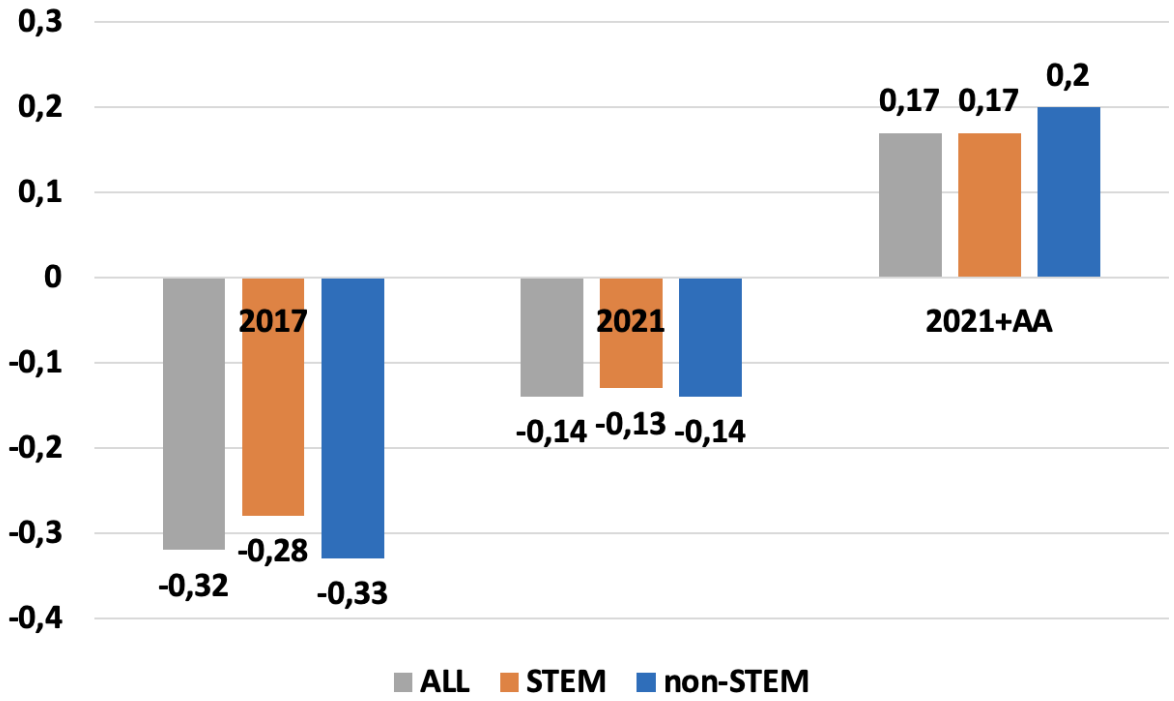
After summing up the additional points from the AA policy, the gender gap is reversed, with female-led groups receiving higher average scores than male-led groups. The comparison between 2017 and 2021 suggests that, even before the AA points were applied, the score gap had already decreased significantly. Furthermore, the affirmative action points fully closed the gender gap in scores.

AA policy and the dynamics of research groups

As seen in the previous section, the data show an increase in the share of applications and the share of groups led by women that are granted funding between 2017 and 2021. We have also documented that the introduction of the AA policy closes the gender gaps in scores relative to the previous calls. In this section, we go one step further and explore the extent to which the composition of the research groups applying to the program is modified once the AA policy is introduced.

Groups that applied for funding were asked to provide information on their participation in previous calls. We use the identifiers reported by the PIs to link groups across calls. We then use these matches to document the dynamic origins of the increase in the share of new female PIs after the introduction of the AA policy in 2021. Table 2 below shows groups that applied for the 2017 and 2021 funding scheme, according to the gender

Figure 1: Gender gaps in scores



Note: The figure shows the difference between the average score received by female-led groups and male-led groups in different years for all fields (grey) and separately for STEM (orange) and non-STEM fields (blue). The gaps are shown for 2017, and for 2021 before and after adding up the additional AA points (2021+AA).

Table 1: Descriptives

	ALL			STEM			Non-STEM		
	2014	**2017**	**2021**	**2014**	**2017**	**2021**	**2014**	**2017**	**2021**
Share of female researchers	45.68%	46.36%	47.78%	44.60%	46.46%	46.49%	48.31%	46.05%	51.61%
Share of groups led by females	27.36%	28.80%	41.35%	22.26%	25.38%	39.58%	39.67%	39.63%	46.58%
Share of granted groups	41.53%	49.20%	63.29%	41.87%	48.75%	63.37%	40.70%	50.61%	36.04%
Share of granted groups led by females	24.05%	24.52%	43.60%	18.61%	20.32%	41.53%	37.56%	37.35%	49.75%
Average score female PI	5.79	7.58	7.67	5.84	7.67	7.71	5.72	7.39	7.56
	[0.79]	[1.33]	[0.89]	[0.72]	[1.25]	[0.90]	[0.86]	[1.46]	[0.84]
Average score male PI	5.87	7.90	7.81	5.93	7.95	7.78	5.67	7.72	7.69
	[0.79]	[1.26]	[0.87]	[0.74]	[1.23]	[0.86]	[0.92]	[1.37]	[0.89]
Average score female PI, with AA			8.12			8.15			8.04
			[0.89]			[0.90]			[0.85]
Average score male PI, with AA			7.95			7.98			7.84
			[0.89]			[0.88]			[0.89]
Number of groups	1652	1368	1272	1168	1040	950	484	328	322

Note: The table is constructed from the administrative data provided by the AGAUR for the three different calls.

of the PI in both years [1]. The upper panel of Table 2 focuses on stable groups, those that

[1] The same table comparing the 2014 and the 2017 calls is included in the Appendix (Table A1).

applied for both calls, in 2017 and 2021. The middle panel focuses on groups that only applied to the 2017 call, while the lower panel focuses on new groups that only applied to the 2021 call. The destruction rate of groups in 2017 is 30%, which means that 440 groups (out of 1430) apply to the 2017 but not to the 2021 call [2]. The creation rate in 2021 is 22%, that is 282 out of 1272 groups are only applying to the 2021 call.

Women represented 28% of those PIs applying to the 2017 call, a share that increased to 41% in 2021. According to the descriptive statistics presented in Table 2, the higher share of female PIs in 2021 is driven by two forces: 1) Among groups applying to both the 2017 and 2021 calls, 19% change from a male PI in 2017 to a female PI in 2021; 2) The share of women-led groups is lower (30%) among groups only applying to the 2017 call than among groups only applying to the 2021 call (43%). Therefore, the increase in the share of female-led groups applying to the 2021 call originates largely from groups that changed from a male to a female PI (13% of the groups in 2021) and, to a lesser extent, from the creation of new groups led by women (9% of the groups in 2021).

Table 2 also reports the average scores received by different types of groups in 2017 and 2021, before and after adding the additional points of the affirmative action policy. If we focus on stable groups (upper panel), in 2017 male-led groups received a higher score than those led by females. The average difference in scores was 0.32 and statistically significant at any conventional level. The table also shows that groups that changed from a male to a female PI between 2017 and 2021, received an almost identical score in 2017 (8.26) than those with a male PI in both calls (8.27). However, if we focus on the scores in 2021 before adding the AA points, we can see that average scores among groups led by male PIs in both calls was 0.11 points higher than among groups that change from a male to a female PI. This difference can be explained by changes in other characteristics of the group or by the existence of a gender bias among reviewers. When we add the score of the AA policy (last column in the table), we see that the policy reverted the magnitude of the gender gap in scores in favor of female PIs. In the following section, we will develop the analysis of the selection process, which will allow us to provide additional insights to this result.

The share of stable groups led by females in both calls represents 27% of the stable groups. These women received the lowest average scores in 2017 but scored the highest in

[2] Notice that in Table 2 the number of groups in 2017 is 1430, higher than in Table 1, where is 1368. This is because the table of group transitions is constructed from the group identifiers in 2017, as reported by groups in 2021. In this setting, 62 groups in 2021 identify with the same group in 2017. For example, group 1047 and 276 in 2021 identify themselves with group 63 in 2017. This can also be explained by the fact that some groups merge in 2021. In this case, the group reports an EXP relative to the group in 2017 and there is no problem.

2021 once we add the AA points. Also, a tiny percentage of stable groups (4%) changed from female to male PI between 2017 and 2021. This group of defiers scored among the highest in 2021 (before the implementation of the AA policy), which points to the promotion of male PIs that were of really high quality (for instance, those that had received a Distinguished Professor Fellow). An interesting observation is that groups that only applied to the 2017 or to the 2021 calls score lower, on average, than stable groups. This is particularly the case for women-led groups before adding the AA points.

Table A1 in the Appendix section shows that these patterns hold for both STEM and non-STEM fields. Figure 2 compares group transitions between the 2014 and 2017 call (grey) and the 2017 and 2021 call (blue). Panel A reports the share of female PIs among groups that are only applying in one of the calls. The label destructed groups refers to groups applying to the 2014 and not to the 2017 call (grey color), as well as those applying to the 2017 but not to the 2021 call (blue color). On the other hand, created groups are those not applying to the 2014 but applying to the 2017 call (grey color), as well as those not applying to the 2017 but applying to the 2021 call (blue color). As it can be seen in the two grey color bars of the figure, between the 2014 and 2017 calls, the share of females was equal among groups that were only applying to one of these two calls. In contrast, the share of female PIs is 13 percentage points higher among groups applying only to the 2021 call than among those applying only to the 2017 call (blue color bars). As discussed, the higher rate of female PIs among newly created groups in 2021 explains part of the increase in the share of female PI under the AA policy (Table A2 in the Appendix reports group transitions between 2014 and 2017).

Panel B focuses only on stable groups between two consecutive calls, which represent 88% and 78% of the groups in the 2014 to 2017 and 2017 to 2021 calls. Among these stable groups, between 2014 and 2017 only 6% of the stable groups changed from a male to a female PI, while this percentage is 13 percentage points higher between the 2017 and 2021 calls (i.e. 19 percentage points). As can be seen in the figure, the other types of transitions according to the gender of the PI are very stable between the 2014-2017 and the 2017-2021 calls.

The analysis of group transitions in Table 2 and Figure 2 is useful to understand the dynamic origins of the increase in female PI applicants when the AA policy is implemented in 2021. This analysis reveals that the larger proportion of female PIs in 2021 responds, on the one hand, to high-scoring 2017 male led groups that changed to female PIs (13%). On the other hand, we also observe that the share of female PIs among new groups in 2021 is higher than among those groups that only applied for the 2017 call. The share of new groups led by females that only apply to the 2021 represent 9% of the

total in this call.

AA policy and changes in the composition of the groups

Our administrative data contains detailed information on all the items included in the application package. Thus, we have the same information than the one used by the external reviewers and the evaluation panels to evaluate and grade all the proposals. As explained above, the evaluation process takes into account both the characteristics of the principal investigator, as well as the characteristics of the groups to decide on the scores. Table A3 shows descriptive statistics on the items used to perform the evaluations. While the structure of the data is similar for 2017 and 2021, there are significant differences in the information collected for the 2014 call. As a result, in what follows we restrict the analysis to the two most recent calls.

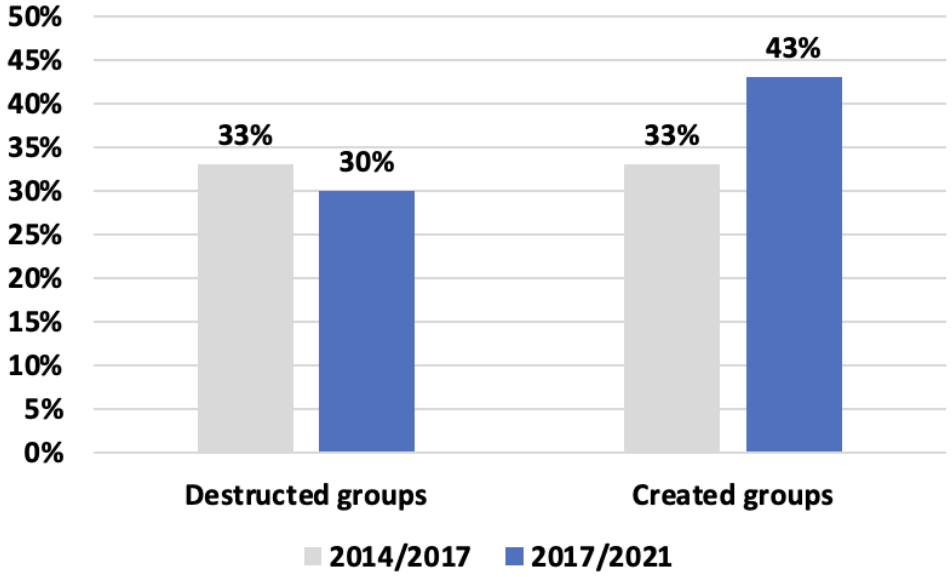
As can be seen in Table A3, groups led by females are different from those led by men. For instance, female PIs are younger and less likely to be full (or Distinguised Full) professors [3]. In terms of the composition of the group, the most relevant differences are that female led groups are smaller in size, concentrate a higher proportion of female researchers and accumulate less publications. Using the information available in the application that contains a ten-line summary of the research developed in the group, we use machine learning techniques to create two indicators that capture the extent to which the research has a gender perspective. The first indicator equals one if the research summary includes at least one of the gender perspective words listed in the Oxford Dictionary for Gender Studies. We construct the second indicator using artificial intelligence tools (more specifically, large language models). We give the model the official definition of gender perspective and train it to identify whether the abstract includes the idea embedded in the official definition of "gender perspective" definition. We can see in Table A3 that, according to our two gender perspective indicators, groups led by women are more likely to incorporate the gender perspective in their research agenda.

Next, we examine the extent to which these differences in PI and group characteristics are statistically significant and are related to the implementation of the AA policy. Table 3 displays the estimated coefficients corresponding to a set of regressions where the dependent variable is a specific characteristic of the PI (or the group). As explanatory variables, the regressions include a gender dummy, an indicator for the year 2021, as well as an interaction term between the gender and the 2021 dummy.

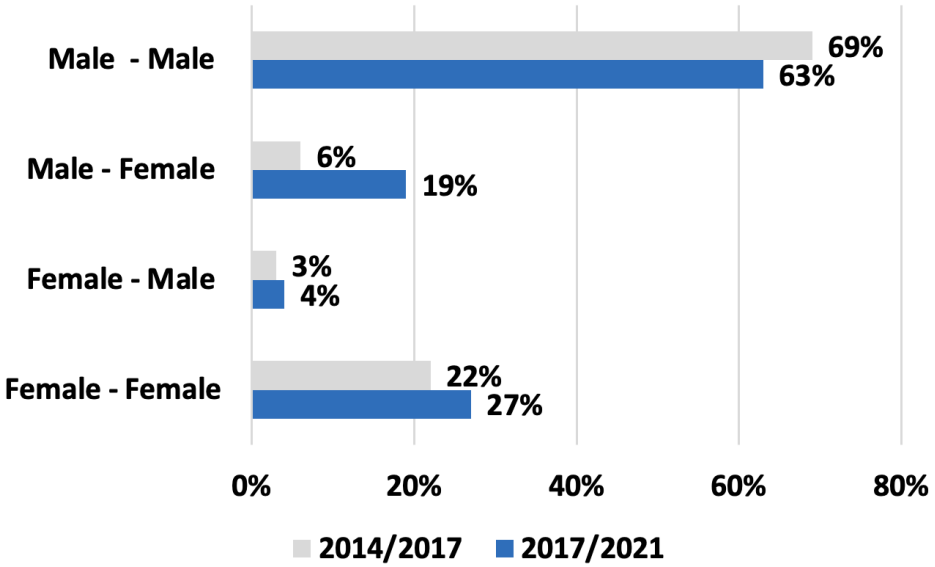
[3] A distinction granted by the Catalan government to the most talented researchers in the region.

Figure 2: Group transitions

Panel A: Share of female PI among destructed and created groups



Panel B: Share of groups by gender of PI in stable groups



Note: The figure compares group transitions between the 2014 and 2017 call (grey) and the 2017 and 2021 call (blue). Panel A displays the share of female PIs among groups that are only applying to one of the calls. If they only apply to the 2014 (grey color) or only to the 2017 call (blue color) they are labelled as destructed groups. If they only apply to the 2017 (grey color) or the 2021 call (blue color) they are labelled as created groups. Panel B restricts the sample to stable groups between two consecutive calls, either between 2014 and 2017 (grey color) or between 2017 and 2021 (blue color). In this panel, the percentages show the share of PIs, according to their gender, in each of these two consecutive calls.

Table 2: Group transitions between the 2017 and 2021 calls

	Female PI 2017	Female PI 2021	Nobs	Percentage	Score 2017	Score 2021	Score 2021 + AA
Stable groups 2017 & 2021	0	0	553	63%	8.26 [1.05]	7.92 [0.79]	8.06 [0.79]
	0	1	166	19%	8.27 [0.95]	7.81 [0.74]	8.23 [0.74]
	1	1	240	27%	7.93 [1.10]	7.80 [0.79]	8.26 [0.80]
	1	0	31	4%	8.01 [0.96]	7.92 [0.63]	8.10 [0.64]
Total			990	100%			
Groups applying only in 2017	0		310	70%	7.22 [1.44]		
	1		130	30%	6.86 [1.47]		
Total			440	100%			
Groups applying only in 2021		0	162	57%		7.40 [1.06]	7.53 [1.07]
		1	120	43%		7.20 [1.08]	7.68 [1.09]
Total			282	100%			

Note: ...

Consistent with what we observed in the descriptive statistics section, Table 3 shows that female PIs are statistically younger and less likely to be full professors than their male counterparts, and these differences persist in 2021. Additionally, female-led groups are smaller in size (less members), are more likely to fulfill the requirement of a 40% female share, accumulate less publications, research recognitions (which are granted by the Catalan Government to individuals with a proven and continuous history of publications during a six-year period), less competitive projects, and are more likely to include the gender perspective in their research. Once the AA policy is implemented, some of these gender differences are reduced. In 2021, groups led by women are larger in size and accumulate more research recognitions and competitive projects, so that the previously existing gender gap in these dimensions narrows down. We also observe a negative effect of including the gender perspective in 2021, which is explained by a higher usage of the gender perspective by male PIs in this last call.

AA policy and changes in the evaluation process

As explained above, in the first step of the selection process, each group is evaluated by three external reviewers (selected from a large pool of potential reviewers). Among all assigned reviewers, there is a 40% gender quota imposed in the pool of reviewers in each call. However, as reviewers are randomly assigned to projects (within field), the exact number of male and female reviewers effectively assigned to each particular project will randomly vary. Once all proposals have been evaluated in this first step, in the second step the evaluation panels rank all the proposals in their field and decide on the threshold to get funded.

Columns (1) and (2) of Table 4 display the raw gender gaps on average scores in 2017 and 2021. As previously shown in Figure 2, in the 2017 call groups led by women received lower scores than those led by men (the difference is 0.31). In 2021, the gap narrows by 50% even before adding the additional points granted by the AA policy. This reduction in the gender gap in scores can be explained by several potential factors, such as a signaling effect of the policy, or changes in the composition of the groups. Column (3) estimates the model with an interaction term between the gender and time dummies. In 2017 groups led by women received on average 0.31 points less in the remote evaluation phase than those led by men, and this gender gap is reduced to 0.14 in 2021, even before taking into account the AA points.

Column (4) takes into account the effect of the gender composition of the evalua-

Table 3: Group Composition: The Effect of Female PIs and Changes Under the Policy

	Female PI	**2021**	**Female PI x 2021**	**Constant**
About the PI				
Full professor	-0.18***	0.03	0.03	0.46***
	[0.03]	[0.02]	[0.04]	[0.02]
Age	-0.91*	-0.34	-0.34	53.31***
	[0.50]	[0.40]	[0.68]	[0.27]
Distinguished FP	-0.04**	0.03**	-0.02	0.10***
	[0.02]	[0.01]	[0.02]	[0.01]
About the Group				
Number of members	-2.15***	1.19***	1.93***	15.99***
	[0.53]	[0.43]	[0.73]	[0.29]
Share of females	0.17***	0.00	-0.03*	0.41***
	[0.01]	[0.01]	[0.01]	[0.01]
Share of females >40%	0.27***	0.01	-0.03	0.54***
	[0.02]	[0.02]	[0.02]	[0.01]
Share with PhD	0.00	0.04***	0.01	0.59***
	[0.01]	[0.01]	[0.01]	[0.00]
Share Senior Postdoc	0.00	0.02***	-0.01	0.02***
	[0.00]	[0.00]	[0.00]	[0.00]
Share Junior Postdoc	-0.01	0.01**	-0.01	0.07***
	[0.01]	[0.01]	[0.01]	[0.00]
Publications	-10.07**	54.19***	-1.46	72.72***
	[4.87]	[3.97]	[6.71]	[2.61]
Research recognition	-1.00**	-7.82***	1.74**	11.38***
	[0.52]	[0.43]	[0.72]	[0.28]
Number of theses defended	-0.15	1.90***	-0.06	7.72***
	[0.35]	[0.29]	[0.49]	[0.19]
Number of patents	-0.22	-0.54	-0.22	1.43**
	[0.35]	[0.29]	[0.49]	[0.19]
Number of competitive projects	-1.69***	15.33***	-1.36	8.21***
	[0.44]	[0.36]	[0.61]	[0.24]
Gender Perspective Oxford Dictionary	0.07**	-0.06***	-0.06*	0.18***
	[0.03]	[0.02]	[0.02]	[0.01]
Gender Perspective Large Language Model	0.04**	0.02***	0.02	0.06***
	[0.02]	[0.01]	[0.02]	[0.01]

Note: Dependent variables are shown in rows. There are three regressions where the dependent variable is a characteristic of the PÎ and thirteen regressions where the dependent variable is a characteristic of the group. The sample is restricted to include only the calls of 2017 and 2021. Explanatory variables are shown in columns.
Standard errors in brackets. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

tors, as well as the interaction term with the gender of the PI. None of the additional regressor is statistically significant. Column (5) controls for the characteristics of the PI (age and full or distinguished full professor). Including these controls does not alter the estimated results and it does not modify the magnitude of the estimated gender gaps.

Column (6) includes the characteristics of the group. When including these additional controls, the gender gap at baseline (in 2017) is reduced by half, and the gender premium for women in 2021 disappears. This effect is mainly driven by the inclusion of the variable

size of the group as a control. As discussed in the previous section, groups led by women are smaller on average, but the size of women-led groups increases in 2021. In column (6) we also display the coefficient on having at least 40% of women in the group, but it is not statistically significant, as most groups were already fulfilling this condition before the AA policy was introduced. Column (7) adds field of study fixed effects, and the results are unaffected, suggesting that the patterns reported in column (6) are present across all fields.

Column (8) displays the effect of adding the AA points. We can see that the AA policy compensated pre-existing unexplained penalties against groups led by women. After controlling for observable characteristics, groups led by women received lower average scores (0.20), and the policy eliminated this differential. In 2021 groups led by women receive on average 0.04 higher scores than their male counterparts. Notice that this last column compares groups with the same observable characteristics (included as controls), with the only difference being the gender of the PI.

Finally, column (9) displays the results for the probability of being selected for funding. From the estimated coefficient, we see that groups led by women had a 12 percentage points lower probability of receiving funding than their male counterparts. With the implementation of the AA policy, the penalty disappears and female-led groups are as likely as male-led groups to receive funding.

From this exercise, we can conclude that the AA policy had significant effects on the promotion of female talent. First, it increased the share of groups led by women that applied for funding. Second, it increased the size and quality (in terms of research recognition) of groups led by women. Finally, the AA policy erased the unexplained scoring penalties against women in the remote evaluation process. As a result, the probability of receiving funding among groups led by women increased with the policy, and it equalized that of male led groups of similar quality. These findings suggest that the AA policy had substantial redistributive effects without compromising efficiency.

Finally, exploiting the random assignment of the three reviewers assigned to each proposal in the remote evaluation phase, in the last column of Table 4 we can see that the share of female reviewers would have to increase from 0 to 50% to have the same effect on the promotion of female talent than the Affirmative Action policy. Given the cost that gender quotas impose on female researchers that act as reviewers (because, as a minority, they are more likely to participate in more committees than their male counterparts), and the limited effects on the promotion of female researchers associated to these policies, we conclude that the AA policy is more cost-effective in promoting female talent.

Conclusions

This paper analyzes the impact of an affirmative action (AA) policy on the promotion of female talent in research groups, focusing on changes in group composition and the evaluation process. We have three waves of the same call for groups recognition and funding in Catalonia, including two pre-reform calls in 2014 and 2017 and one post-reform call in 2021. Our dataset includes all the information contained in the application package, so that we observe the same quality measures than the reviewers of the program. Our results provide compelling evidence of the policy's effectiveness in addressing gender disparities, promoting female leadership in research, and improving the evaluation outcomes for women-led groups.

More specifically, we document four important findings in the paper. First, the implementation of the AA policy led to a significant increase in the share of female-led research groups. This change was driven by two key factors: (1) A notable shift in the gender of the principal investigator (PI) in existing groups, with 19% of the groups in 2021 transitioning from male to female PI, and (2) The creation of new groups led by women, which account for 9% of the groups in 2021. This shift reflects the successful promotion of female leadership in research as a result of the policy.

Second, female-led groups in 2021 were larger and accumulated more research recognition elements compared to their counterparts in 2017, narrowing the gender gap in these dimensions. However, women-led groups still lagged behind in terms of the total number of publications suggesting that, while progress has been made, there is still room for further improvement in the academic output of female-led groups.

Third, prior to the introduction of the AA policy, groups led by women consistently received lower evaluations scores than those led by men, with an average gap of 0.31 points in 2017. By 2021, the gap had narrowed by half before taking into account the additional AA points. This reduction largely responds to the larger size of groups led by women in 2021. Finally, when we take into account the additional AA points, the policy effectively eliminates unexplained penalties against women in the evaluation process, leading to more equitable outcomes.

The paper also compares the AA policy to other potential strategies for promoting gender equity in research environments, such as increasing the share of female reviewers in the evaluation process. The analysis suggests that the AA policy is more cost effective,

Table 4: The effect of the policy on the evaluation process and the gender gaps

VARIABLES	**[1]** score	**[2]** score	**(3)** score	**(4)** score	**(5)** score	**(6)** score	**(7)** score	**(8)** score + AA	**(9)** granted
Female PI	-0.32***	-0.15***	-0.32***	-0.34***	-0.28***	-0.13*	-0.13*	-0.17**	-0.12***
	[0.08]	[0.05]	[0.07]	[0.09]	[0.08]	[0.08]	[0.08]	[0.08]	[0.04]
2021			0.32***	0.29***	0.27***	0.18**	0.20***	-0.07	0.00
			[0.05]	[0.08]	[0.08]	[0.08]	[0.08]	[0.08]	[0.04]
Female PI × 2021			0.18*	0.20	0.23*	-0.03	-0.05	0.27**	0.15***
			[0.09]	[0.14]	[0.14]	[0.12]	[0.12]	[0.12]	[0.06]
Share female reviewers				-0.01	-0.01	-0.06	0.11	0.09	-0.03
				[0.12]	[0.12]	[0.10]	[0.10]	[0.10]	[0.05]
Share fem rev × 2021				0.08	0.06	-0.14	-0.17	-0.14	-0.05
				[0.18]	[0.18]	[0.16]	[0.16]	[0.16]	[0.07]
Female PI × Share fem rev				0.08	0.13	0.04	0.06	0.07	0.27***
				[0.22]	[0.22]	[0.19]	[0.19]	[0.19]	[0.09]
Female PI × Share fem rev × 2021				-0.10	-0.14	0.23	0.25	0.24	-0.14
				[0.31]	[0.30]	[0.26]	[0.26]	[0.26]	[0.12]
Share of females 40%						-0.01	-0.05	0.07	0.02
						[0.04]	[0.05]	[0.05]	[0.02]
Constant	7.90***	8.22***	7.90***	7.91***	7.51***	7.02***	6.99***	6.93***	0.57***
	[0.04]	[0.03]	[0.04]	[0.05]	[0.12]	[0.14]	[0.15]	[0.14]	[0.07]
Year call	2017	2021	2017&2021	2017&2021	2017&2021	2017&2021	2017&2021	2017&2021	2017&2021
PI characteristics	NO	NO	NO	NO	YES	YES	YES	YES	YES
Research group characteristics	NO	NO	NO	NO	NO	YES	YES	YES	YES
Field of study FE	NO	NO	NO	NO	NO	NO	YES	YES	YES
Observations	1,367	1,272	2,639	2,639	2,639	2,639	2,639	2,639	2,639
R-squared	0.01	0.01	0.03	0.03	0.10	0.30	0.30	0.29	0.21

Note: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

as it directly compensates for gender biases in the evaluation process without imposing significant additional costs on female researchers.

In conclusion, the AA policy has a significant and positive impact on the promotion of female talent in research. Not only increases the representation of women in leadership positions, but it also improves the evaluation outcomes of female-led groups, compensating previous unexplained penalties. These findings underscore the importance of targeted interventions in addressing gender inequalities in academic research settings. While challenges remain, particularly in areas like the likelihood of getting published or cited, the AA policy represents an effective and efficient approach to promoting gender equality in science. Future research should explore the long-term effects of such policies.

References

Alexander, D. et al. (2023). *Gender and the time cost of peer review*. Tech. rep. TI 2023-044/V. Tinbergen Institute Discussion Paper.

Arcidiacono, Peter and Michael Lovenheim (2016). "Affirmative action and the quality–fit trade-off". In: *Journal of Economic Literature* 54.1, pp. 3–51.

Ayarza, Alaitz and Nagore Iriberri (2024). "Gender Gaps among Scholars in Economics: Analysis Across Cohorts".

Babcock, L. et al. (2017). "Gender differences in accepting and receiving requests for tasks with low promotability". In: *American Economic Review* 107.3, pp. 714–747.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva (2017). "Does the gender composition of scientific committees matter?" In: *American Economic Review* 107.4, pp. 1207–1238.

Bagues, Manuel F. and Berta Esteve-Volart (2010). "Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment". In: *The Review of Economic Studies* 77.4, pp. 1301–1328.

Balafoutas, Loukas and Matthias Sutter (2012). "Affirmative action policies promote women and do not harm efficiency in the laboratory". In: *Science* 335.6068, pp. 579–582.

Baltrunaite, Audinga, Alessandra Casarico, and Lucia Rizzica (2024). *Women in Economics: The role of gendered references at entry in the profession*. Tech. rep. 1438. Bank of Italy Temi di Discussione (Working Paper).

Bateman, Victoria and Erin Hengel (2023). "The gender gap in UK academic economics 1996-2018: progress, stagnation and retreat". In: *Œconomia. History, Methodology, Philosophy* 13.2, pp. 163–200.

Beaman, L. et al. (2009). "Powerful women: does exposure reduce bias?" In: *The Quarterly Journal of Economics* 124.4, pp. 1497–1540.

Bleemer, Zachary (2022). "Affirmative action, mismatch, and economic mobility after California's Proposition 209". In: *The Quarterly Journal of Economics* 137.1, pp. 115–160.

Boring, Anne (2017). "Gender biases in student evaluations of teaching". In: *Journal of Public Economics* 145, pp. 27–41.

Buckles, Kasey (2019). "Fixing the leaky pipeline: Strategies for making economics work for women at every stage". In: *Journal of Economic Perspectives* 33.1, pp. 43–60.

Canadian Association of University Teachers (2020). *Equity Report 2020*.

Card, D. et al. (2020). "Are referees and editors in economics gender neutral?" In: *The Quarterly Journal of Economics* 135.1, pp. 269–327.

— (2022). "Gender differences in peer recognition by economists". In: *Econometrica* 90.5, pp. 1937–1971.

Card, D. et al. (2023). "Gender gaps at the academies". In: *Proceedings of the National Academy of Sciences* 120.4, e2212421120.

Chari, Anusha and Paul Goldsmith-Pinkham (2017). *Gender representation in economics across topics and time: Evidence from the NBER summer institute*. Tech. rep. w23953. National Bureau of Economic Research.

Dupas, P. et al. (2021). *Gender and the dynamics of economics seminars*. Tech. rep. w28494. National Bureau of Economic Research.

Eberhardt, Markus, Giovanni Facchini, and Valeria Rueda (2023). "Gender differences in reference letters: Evidence from the economics job market". In: *The Economic Journal* 133.655, pp. 2676–2708.

European Commission (2021). *She Figures 2021: Gender in Research and Innovation*. Publications Office of the European Union. DOI: 10.2777/06090.

Farré, Lídia and Francesc Ortega (2023). "Selecting Talent: Gender Differences in Success in Competitive Selection Processes". In: *Journal of Human Resources* 58.6, pp. 1881–1913.

Funk, Patricia, Nagore Iriberri, and Nicole Venus (2024). *Women in Editorial Boards: An Investigation of Female Representation in Top Economic Journals*. Tech. rep. 19303. CEPR Discussion Papers.

Gamage, Danula K. and Almudena Sevilla (2019). "Gender equality and positive action: evidence from UK universities". In: *AEA Papers and Proceedings*. Vol. 109. American Economic Association.

Ginther, Donna K. et al. (2020). "Can mentoring help female assistant professors in economics? An evaluation by randomized trial". In: *AEA Papers and Proceedings*. Vol. 110. American Economic Association.

Ginther, Donna K. and Shulamit Kahn (2004). "Women in economics: moving up or falling off the academic career ladder?" In: *Journal of Economic Perspectives* 18.3, pp. 193–214.

Handlan, A. and H. Sheng (2023). *Gender and tone in recorded economics presentations: Audio analysis with machine learning*.

Hengel, Erin (2022). "Publishing while female: Are women held to higher standards? Evidence from peer review". In: *The Economic Journal* 132.648, pp. 2951–2991.

Holzer, Harry and David Neumark (2000). "Assessing affirmative action". In: *Journal of Economic Literature* 38.3, pp. 483–568.

Hospido, Laura and Carlos Sanz (2021). "Gender gaps in the evaluation of research: Evidence from submissions to economics conferences". In: *Oxford Bulletin of Economics and Statistics* 83.3, pp. 590–618.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger (2024). *Gender gaps in academia: Global evidence over the twentieth century*.

Kahn, Shulamit (1995). "Women in the economics profession". In: *Journal of Economic Perspectives* 9.4, pp. 193–205.

Koffi, Marlène (2021). "Gendered citations at top economic journals". In: *AEA Papers and Proceedings*. Vol. 111. American Economic Association.

Lundberg, Shelly and Jenna Stearns (2019). "Women in economics: Stalled progress". In: *Journal of Economic Perspectives* 33.1, pp. 3–22.

Mello, Ursula (2023). "Affirmative action and the choice of schools". In: *Journal of Public Economics* 219, p. 104824.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz (2019). "Gender bias in teaching evaluations". In: *Journal of the European Economic Association* 17.2, pp. 535–566.

National Center for Education Statistics (2022). *Condition of Education 2022*. U.S. Department of Education.

Sarsons, H. et al. (2021). "Gender differences in recognition for group work". In: *Journal of Political Economy* 129.1, pp. 101–147.

Seré, Mateo (2023). *Don't Stop Me Now: Gender Attitudes in Academic Seminars Through Machine Learning*. Tech. rep. 2309.

Tomkins, Andrew, Min Zhang, and William D. Heavlin (2017). "Reviewer bias in single- versus double-blind peer review". In: *Proceedings of the National Academy of Sciences* 114.48, pp. 12708–12713.

Venus, Nicole (2024). *The representation of female economists on Wikipedia*.

Appendix

Table A1: Group transition between 2017 and 2021 in STEM and non-STEM fields

STEM

Group	**Female PI**		**Nobs**	**Percentage**	**Score**		
	2017	2021			2017	2021	2021 + AA
Stable groups							
(2017 & 2021)	0	0	437	58%	8.31 [1.00]	7.96 [0.77]	8.1 [0.79]
	0	1	141	19%	8.26 [0.90]	7.82 [0.74]	8.23 [0.74]
	1	1	164	22%	7.98 [1.09]	7.85 [0.80]	8.31 [0.80]
	1	0	16	2%	8.35 [0.74]	8.1 [0.59]	8.25 [0.62]
Total			758				
Exiting groups							
(only 2017)	0	.	247	74%	7.28 [1.44]		
	1	.	89	26%	7.00 [1.31]		
Total			336				
Entering groups							
(only 2021)	.	0	122	61%		7.44 [1.05]	7.55 [1.06]
	.	1	78	39%		7.17 [1.11]	7.64 [1.13]
Total			200				

Non-STEM

Group	**Female PI**		**Nobs**	**Percentage**	**Score**		
	2017	2021			2017	2021	2021 + AA
Stable groups							
(2017 & 2021)	0	0	116	50%	8.06 [1.19]	7.77 [0.81]	7.92 [0.81]
	0	1	25	11%	8.30 [1.20]	7.77 [0.76]	8.22 [0.76]
	1	1	76	33%	7.81 [1.12]	7.67 [0.77]	8.14 [0.78]
	1	0	15	6%	7.65 [1.05]	7.73 [0.64]	7.93 [0.65]
Total			232				
Exiting groups							
(only 2017)	0	.	63	61%	7.00 [1.44]		
	1	.	41	39%	6.56 [1.76]		
Total			104				
Entering groups							
(only 2021)	.	0	40	49%		7.31 [1.10]	7.46 [1.10]
	.	1	42	51%		7.29 [1.01]	7.77 [1.00]
Total			82				

Figure A1: Share of female researchers and share of female PIs

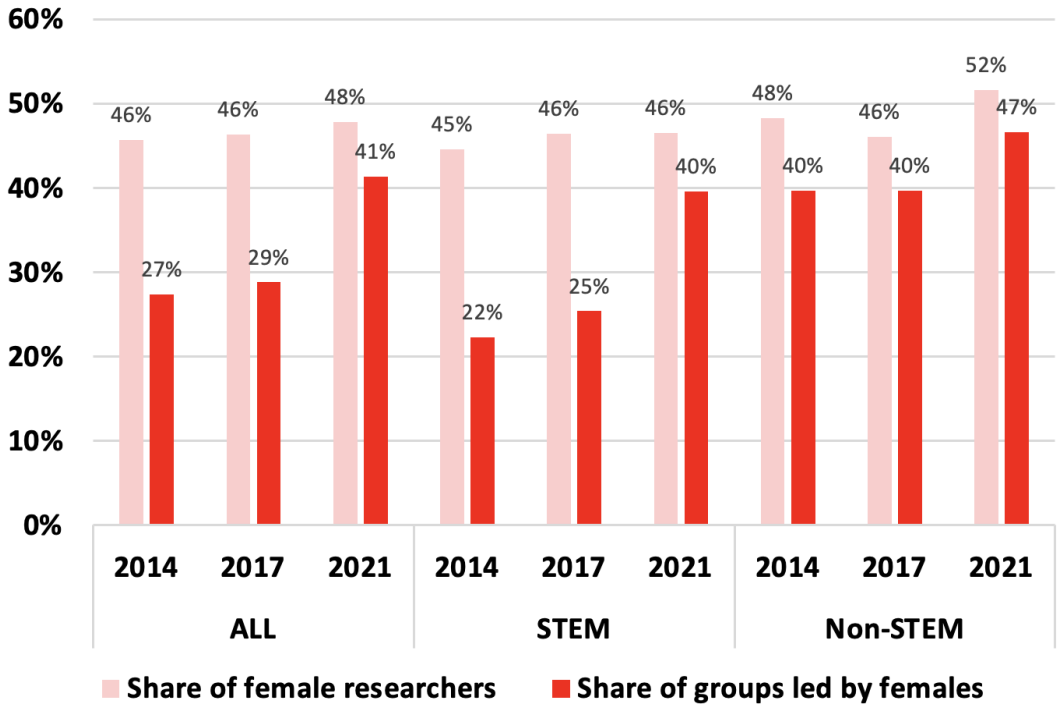


Figure A2: Share of granted groups and share of granted groups led by females

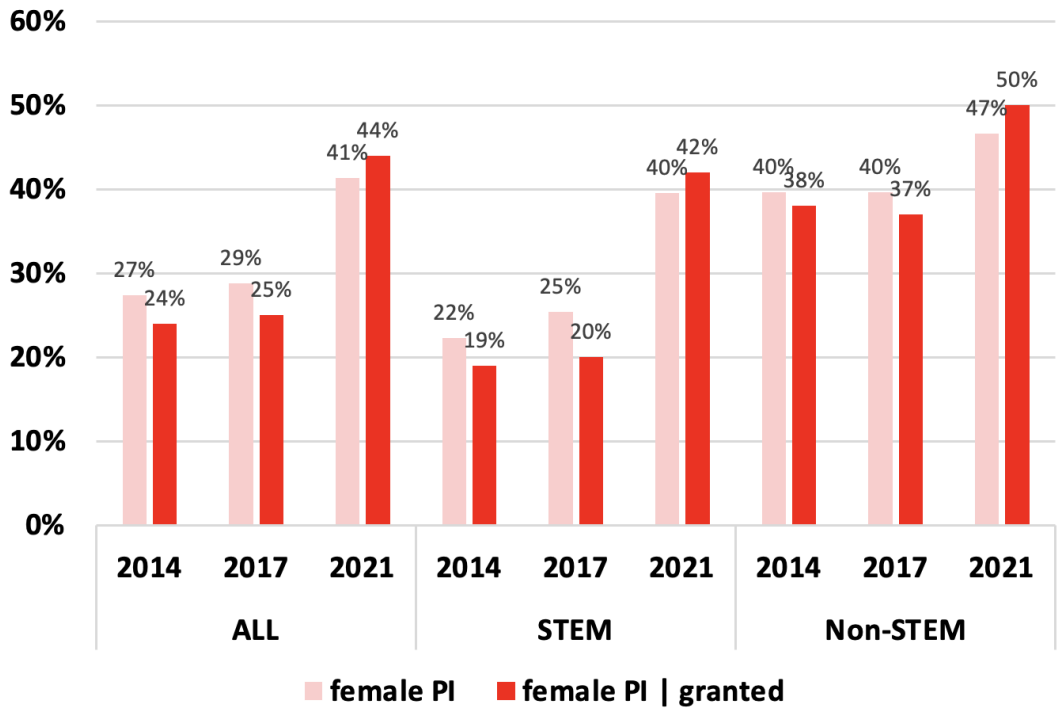


Table A2: Group transitions between 2014 and 2017

	Female PI		**Nobs**	**Percentage**	**Score**	
Group	2014	2017			2014	2017
Stable groups						
(2014 & 2017)	0	0	832	69%	6.09 [0.62]	8.04 [1.50]
	0	1	74	6%	6.14 [0.47]	7.72 [1.29]
	1	1	267	22%	6.04 [0.62]	7.70 [1.21]
	1	0	36	3%	5.93 [0.59]	7.56 [1.26]
Total			1209			
Exiting groups						
(only 2014)	0	.	327	67%	5.28 [0.91]	
	1	.	158	33%	5.35 [0.87]	
Total			485			
Entering groups						
(only 2017)	.	0	106	67%		6.94 [1.66]
	.	1	53	33%		6.77 [1.62]
Total			159			

Table A3: Descriptive statistics by groups

	2014		2017		2021	
	Female	Male	Female	Male	Female	Male
About the PI						
Share of Full Professor	27,21%	45,75%	27,92%	45,79%	34,03%	49,06%
Age	53,21	53,28	52,4	53,3	51,71	52,97
Distinguished Full Professor (ICREA)	3,10%	7,33%	5,08%	9,55%	5,70%	12,47%
About the group						
Number of members	12,23	13,71	13,84	15,98	16,95	17,17
Share of females	61,09%	39,88%	58,78%	41,34%	56,36%	41,74%
Share of females >40%	83,85%	52,17%	81,47%	54,11%	79,85%	55,50%
Share of members with PhD			58,34%	58,61%	63,18%	62,40%
Share of Distinguished Full Professor (ICREA)			1,18%	2,15%	1,76%	2,22%
Share Senior Postdocs Talent Recognition			1,96%	1,52%	3,09%	3,32%
Share Junior Postdocs Talent Recognition			5,91%	6,74%	7,75%	7,78%
Research recognition (Sexennis)			10,38	11,37	4,3	3,56
Number of publications			62,65	72,64	115,38	126,91
Number of thesis defended	6,47	7,35	6,97	7,12	8,82	9,02
Number of patents			0,87	1,43	1,2	1,97
Number of competitive projects			6,52	8,21	9,84	10,14
Gender Perspective Dictionary	21,68%	12,83%	25,38%	18,28%	25,48%	24,66%
Gender Perspective Large Language Models	9,73%	4,75%	10,15%	6,47%	14,07%	8,71%
STEM	57,52%	75,67%	67,00%	79,67%	71,48%	76,94%
Number of groups	452	1200	394	974	526	746

Table A4: The evaluation process ALL, STEM, and Non-STEM

	ALL			STEM			Non-STEM		
VARIABLES	**(1) score**	**(2) score + AA**	**(3) granted**	**(4) score**	**(5) score + AA**	**(6) granted**	**(7) score**	**(8) score + AA**	**(9) granted**
Female PI	-0.16**	-0.20***	-0.13***	-0.16*	-0.19**	-0.14***	-0.16	-0.22	-0.05
	[0.08]	[0.08]	[0.04]	[0.09]	[0.08]	[0.04]	[0.19]	[0.18]	[0.08]
2021	0.20***	-0.07	0.01	0.18**	-0.10	0.00	0.34	0.08	0.02
	[0.08]	[0.08]	[0.04]	[0.08]	[0.08]	[0.04]	[0.21]	[0.20]	[0.09]
Female PI × 2021	-0.01	0.31**	0.17***	0.05	0.37***	0.20***	-0.22	0.14	0.05
	[0.13]	[0.12]	[0.06]	[0.14]	[0.13]	[0.06]	[0.32]	[0.31]	[0.14]
Share female reviewers	0.12	0.10	-0.03	0.04	0.02	-0.07	0.29	0.27	0.10
	[0.11]	[0.11]	[0.05]	[0.12]	[0.12]	[0.06]	[0.24]	[0.24]	[0.11]
Share fem rev × 2021	-0.17	-0.14	-0.04	-0.15	-0.12	-0.01	-0.15	-0.12	-0.08
	[0.16]	[0.16]	[0.07]	[0.18]	[0.18]	[0.09]	[0.36]	[0.36]	[0.16]
Female PI × Share fem rev	0.10	0.11	0.28***	0.30	0.31	0.36***	-0.19	-0.17	0.06
	[0.19]	[0.19]	[0.09]	[0.25]	[0.24]	[0.12]	[0.37]	[0.36]	[0.16]
Female PI × Share fem rev × 2021	0.19	0.19	-0.17	-0.15	-0.16	-0.27*	0.74	0.69	0.10
	[0.27]	[0.26]	[0.12]	[0.33]	[0.32]	[0.15]	[0.55]	[0.54]	[0.24]
Share of females 40%	-0.03	0.09**	0.03	0.01	0.13**	0.05*	-0.06	0.06	0.01
	[0.05]	[0.05]	[0.02]	[0.05]	[0.05]	[0.03]	[0.10]	[0.10]	[0.04]
Constant	6.95***	6.89***	0.55***	6.97***	6.92***	0.56***	6.70***	6.67***	0.45***
	[0.15]	[0.15]	[0.07]	[0.16]	[0.16]	[0.08]	[0.36]	[0.36]	[0.16]
Year call	2017	2021	2017&2021	2017&2021	2017&2021	2017&2021	2017&2021	2017&2021	2017&2021
PI characteristics	NO	NO	NO	NO	YES	YES	YES	YES	YES
Research group characteristics	NO	NO	NO	NO	NO	YES	YES	YES	YES
Observations	2,639	2,639	2,639	1,989	1,989	1,989	650	650	650
R-squared	0.27	0.26	0.18	0.28	0.27	0.21	0.25	0.24	0.13

Note: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.