Female Promotions and the Academic Pipeline: Evidence from a Natural Experiment

Manuel Bagues Milan Makany Giulia Vattuone Natalia Zinovyeva*

March 16, 2026

Abstract

We study how faculty promotion decisions shape women's careers and the academic pipeline, using data from 4,000 Spanish university departments across all disciplines. We identify exogenous variation in promotions using the random assignment of evaluators to promotion committees between 2002 and 2008: applicants whose committees included a co-author or colleague were significantly more likely to qualify for promotion. We document two main findings. First, failing to obtain tenure has asymmetrically lasting consequences for women. Those who narrowly miss tenure are 57 percentage points less likely to be tenured fifteen years later, compared to 29 percentage points for men. Second, when women do obtain tenure, the effects extend well beyond their own careers: promoting a woman to Associate Professor increases female faculty by 1.5 members after 15 years, leads to six additional female PhD graduates over the following decade, and raises the number who subsequently remain in academia and reach tenured positions.

JEL Codes: I23, J16, J44, M51

Keywords: Academic Promotions, Women in Academia, Natural Experiment

*Manuel Bagues, University of Warwick, CEPR, IZA and J-Pal; email: manuel.bagues@warwick.ac.uk; Milan Makany, Erasmus University Rotterdam, Tinbergen Institute, email: makany@ese.eur.nl; Giulia Vattuone, SOFI, Stockholm University, email: giulia.vattuone@sofi.su.se; Natalia Zinovyeva, University of Warwick; email: natalia.zinovyeva@warwick.ac.uk. We thank for useful comments Michèle Belot, Anne Boring, Alessandra Casarico, Igo Dobbe, Olle Folke, Christoph Hedtrich, Bas van der Klaauw, Marco Nieddu, Lorenzo Pandolfi, Johanna Rickne, David Strömberg, Ekaterina Zhuravskaya, and participants in presentations at Warwick University, University of Zurich, Paris School of Economics, University College London, University of Exeter, the COSME Workshop on Gender Equality, Birkbeck University of London, CEMFI, University of Oviedo, Stockholm University, Stockholm School of Economics, Erasmus School of Economics, University of Oxford, University of Oslo, BI Norwegian Business School, University of Naples Federico II, Centre d'Économie de la Sorbonne, CY Cergy Paris University, and the Tinbergen Institute.

1 Introduction

Women remain underrepresented in the upper ranks of academia. In the United States and in the European Union, women earn nearly half of all Ph.D. degrees but hold only about 40% of associate professorships and a third of full professorships (European Commission, 2024; U.S. National Science Foundation, 2024; U.S. Department of Education, 2022). The disparities are particularly pronounced in Science, Technology, Engineering, and Mathematics (STEM), where fewer than one-sixth of full professors are women, but even in fields with relatively higher female representation, such as Social Sciences and Humanities, women hold less than 40% of full professorships. Economics presents one of the starkest gender gaps: women account for only 13% of full professors in top-ranked departments in Europe and 15% in the US (Lundberg and Stearns, 2019; Giraneza Birekeraho and Maniga, 2018). Despite decades of growth in women's doctoral attainment, these gaps have proven remarkably persistent, reflecting the so-called "leaky pipeline", a metaphor for the disproportionately high attrition and slower advancement women experience at successive stages of the academic career ladder (Lundberg and Stearns, 2019; Iaria et al., 2022). Beyond their equity implications, these persistent gender gaps represent a misallocation of talent with significant consequences for scientific progress and economic growth (Hsieh et al., 2019).

This continued underrepresentation has prompted a variety of policy interventions designed to increase female representation at senior ranks, often grounded in the hope that greater female presence at the top will generate broader benefits along the academic pipeline. These interventions range from targeted hiring practices to financial incentives. Some universities have introduced women-only professorships and some countries have tied research funding to hiring subsidies for women, as Germany has done since 2012 (Wissenschaftskonferenz, 2016).[1] Additionally, many departments all over the world have adopted voluntary gender balance

1. For instance, women-only full-professor posts have been opened at the University of Adelaide in Australia (https://www.abc.net.au/news/2018-08-22), Eindhoven University of Technology in the Netherlands (https://www.science.org/content/article/men-need-not-apply-university-set-open-jobs-just-women), the Vienna University of Technology in Austria (https://informatics.tuwien.ac.at/news/1429), and across a dozen institutions in Ireland (https://www.siliconrepublic.com/innovation/women-only-professor-positions-sali-ireland).

targets for senior academic hires.

In this paper we investigate whether higher female representation in top academic positions produces benefits that trickle down to women at earlier career stages, including PhD students and junior faculty. There are several theoretical reasons to expect such effects. First, female representation among senior faculty may improve departmental climate, making it more inclusive. Survey evidence suggests this is a pressing concern: in a 2019 American Economic Association survey, 40% of women reported dissatisfaction with the professional climate and 44% did not feel socially included in their field. A stronger presence of senior women can foster more inclusive cultures and help curb sexual harassment (Folke and Rickne, 2022), while also advocating for family-friendly policies that help retain women in academia (Epifanio and Troeger, 2020). Second, senior women may serve as role models and mentors for junior scholars (Ginther et al., 2020; Carrell et al., 2010; Porter and Serra, 2020; Bettinger and Long, 2005).[2] Third, it has been argued that the presence of women in top positions may help eliminate gender discrimination in hiring and promotion, although evidence on whether female evaluators favor female applicants is mixed, with several studies finding no such effect (Broder, 1993; Moss-Racusin et al., 2012; Williams and Ceci, 2015; Bagues et al., 2017; Card et al., 2020; Deschamps, 2024).[3] Finally, research subfields are not gender-neutral: women and men tend to cluster in different areas (Dolado et al., 2012; Huang et al., 2020; Kozlowski et al., 2022). In male-dominated departments, hiring and promotion decisions

2. Ginther et al., 2020 find, using a randomized control trial, that junior female economists mentored by senior women experienced significant improvements in research productivity and job satisfaction. Carrell et al., 2010 show that female students at the U.S. Air Force Academy are 26 percentage points more likely to complete a STEM major when taught by female instructors. Porter and Serra, 2020 find that exposure to successful women increases female economics majors by 8 percentage points. However, Bettinger and Long, 2005 report positive effects in Sociology but negative effects in Economics and other fields.

3. Broder, 1993 examined reviews of NSF proposals showing that female-authored proposals receive lower ratings from female reviewers. Williams and Ceci, 2015 and Moss-Racusin et al., 2012 find that the gender gap is similar among male and female evaluators in several experiments where faculty evaluated hypothetical female and male applicants for assistant professorships and a laboratory manager position respectively. Bagues et al., 2017 analyzed data from applications for qualification as Associate and Full Professor in all academic disciplines in Spain and Italy, finding that female applicants have slightly lower chances of success when they are randomly assigned to a committee with a higher share of female evaluators. Deschamps, 2024 reaches similar conclusions when analyzing the effects of a law-mandated increase in female representation on French academic hiring committees. Card et al., 2020 considered information from submissions to four leading journals in Economics, and they do not observe any significant difference in the evaluations received by female authors depending on the gender of referees.

may systematically disadvantage candidates working in these areas — not necessarily due to gender bias per se, but because evaluators tend to favor research closer to their own agenda (Bagues et al., 2017). More women in senior roles could diversify research priorities and help correct these biases (Bello et al., 2025; Belot et al., 2023).

While these mechanisms suggest potential benefits from increased female representation in senior positions, their actual impact may be limited. Senior women who have successfully navigated male-dominated hierarchies may represent a highly selected subset whose preferences and working styles resemble those of their male colleagues, potentially limiting their capacity or inclination to advocate for systemic change. Even when they do seek change, minority status may prevent them from implementing their views.

Empirically evaluating these theoretical benefits has proven challenging. While there is substantial evidence supporting individual mechanisms – such as the positive effects of female role models through classroom exposure or mentorship relationships – there is remarkably little causal evidence on whether increasing female representation in senior faculty positions actually generates the hypothesized trickle-down effects. The existing literature on departmental gender composition relies primarily on observational correlations, which may reflect omitted variables such as departmental culture or institutional policies that simultaneously attract female faculty and improve outcomes for women (e.g., Boustan and Langan, 2019).

We provide novel evidence on the causal effect of promoting women to top positions, exploiting a unique natural experiment: the centralized qualification system in Spanish academia between 2002 and 2008. During this period, all researchers seeking promotion to Associate or Full Professor had to first qualify in a national evaluation, with committee members selected via lottery from eligible evaluators. Crucially, no conflict of interest rules were in place, so applicants could be randomly assigned to committees that included personal academic connections such as co-authors, colleagues, or thesis advisors. As Zinovyeva and Bagues, 2015 show, such connections significantly increased applicants' chances of qualifying. Since nearly all qualified candidates were subsequently promoted by their departments, this generates quasi-random variation in departmental gender composition at different ranks, which we exploit to estimate the causal effect of female faculty representation on junior women's outcomes. Due to the competitive nature of the *habilitación*, the compliers in our setting tend to have stronger research records than the average candidate, meaning our estimates speak most directly to the effects of promoting highly capable women at the margin.

Our analysis covers all academic disciplines, drawing on data from all public universities in Spain — representing over 90% of the higher education system — and spanning 4,000 departments from 1985 to 2023, including faculty composition by rank and gender, individual research production, and PhD graduate characteristics. We first document that, at the department level, female representation at senior ranks correlates positively with women's presence at lower ranks and in PhD programs, consistent with evidence from Economics in the US (Boustan and Langan, 2019). Female PhD students are also substantially more likely to be advised by female faculty. However, these correlations may reflect unobserved confounders rather than causal effects. We therefore estimate the causal impact of promoting women to senior faculty positions, exploiting quasi-random variation in committee composition generated by the random assignment of evaluators to qualification exams.

To assess the validity of our empirical strategy, we verify that qualification exam outcomes have persistent effects on individual career trajectories. Candidates who qualify due to a favorable committee draw obtain promotions within one to two years, and remain significantly more likely to have been promoted to their target rank even a decade later — despite the fact that initially unsuccessful candidates could, and often did, qualify in later rounds. The effect is particularly large and persistent for Associate Professor positions, where exogenous qualification raises the probability of tenure after ten years by 48 percentage points, with the gap remaining above 40 percentage points even fifteen years later. Strikingly, this persistence is more pronounced for women than for men: ten years after the exam, the promotion gap between lucky and unlucky candidates is 83 percentage points for women and 38 percentage points for men, suggesting that failing to obtain tenure has especially lasting consequences for women's academic careers.

For Full Professor positions, effects are smaller and become statistically insignificant after around six years, likely reflecting that Full Professor candidates already hold a tenured Associate Professor position and can reapply in subsequent rounds; gender-specific estimates at this rank are particularly imprecise given the small number of female Full Professor candidates. The transitory nature of the Full Professor shock limits our ability to identify downstream effects, and we therefore focus on Associate Professor qualifications for the remainder of the analysis.

We then examine how female Associate Professor qualifications affect departmental gender composition. In the short term, an exogenous female promotion translates mechanically into a proportional increase in female representation. Over the longer term, this effect grows: after 15 years, affected departments have approximately 1.5 more female Associate Professors than departments where the female applicant did not qualify, of whom roughly one represents additional women beyond the promoted candidate herself.

We then analyze how female Associate Professor qualifications affect PhD students. Each exogenously promoted woman leads to approximately six additional female PhD graduates and four who remain research-active over a 10-year horizon; these graduates are also significantly more likely to hold a tenured position ten years after graduation. These patterns suggest that sustained female representation at the faculty level increases both the quantity of female researchers entering the pipeline and their persistence within it.[4]

These effects — both on faculty composition and on PhD outcomes — are concentrated in departments where women are present but not yet in the majority, suggesting that trickle-down effects require a critical mass rather than being triggered by the mere promotion of a first woman into an all-male department.

One possible channel through which trickle-down effects may operate is research orientation rather than gender per se. Women are more likely to work in research areas with higher female representation, and these areas may attract more female students and future faculty. To disentangle these channels, we examine whether promoting researchers working in more feminized research topics, irrespective of their gender, produces similar effects. We find that such promotions also increase female representation among faculty, suggesting that hiring in

4. Our findings are consistent with descriptive studies documenting that female PhD students with female advisors tend to be more successful and academically oriented (C. Hilmer and M. Hilmer, 2007; Gaule and Piacentini, 2018; Hofstra et al., 2022; Rossello et al., 2024), though some earlier evidence finds no such effects (Neumark and Gardecki, 1998).

more feminized research areas can help improve gender balance at the faculty level. However, their effects on female PhD graduates are weaker than those generated by promoting a female researcher directly. This asymmetry points to mechanisms specific to gender — such as mentoring, role models, and access to professional networks — as playing an important role in shaping PhD outcomes, above and beyond the indirect channel of research alignment.

This paper makes four main contributions. First, we provide novel causal evidence on the individual cost of failing to obtain tenure, showing that this cost is substantially larger for women than for men: marginally unsuccessful female candidates are significantly less likely to be tenured even ten years later. This gender asymmetry suggests that missing tenure at a critical early stage has lasting consequences for women that it does not have for men, leaving many permanently outside the tenured ranks. These findings point to early tenure failures as an underappreciated driver of persistent female underrepresentation in academia, and parallel evidence from other competitive settings — such as electoral politics — where initial setbacks have been shown to have asymmetrically lasting consequences for women (Wasserman, 2023).

Second, we show that when women do obtain tenure, the effects extend beyond their own career progression to benefit junior women in their departments. We provide the first large-scale causal evidence on how increasing the representation of women in senior academic positions affects junior researchers. Prior work has either relied on observational correlations across departments (Boustan and Langan, 2019; Hofstra et al., 2022), which cannot separate representation effects from underlying culture or selection, or on small-scale interventions with limited generalizability (e.g., Ginther et al., 2020). Exploiting quasi-random variation from Spain's national qualification system, we identify persistent downstream effects of female promotions at the departmental level, showing that additional female Associate Professors lead to sustained increases in the number of female Associate Professors fifteen years later and substantially expand the pipeline of female PhD graduates and future academics.

Third, we contribute to understanding why representation matters in academic settings. By comparing the impact of promoting women with the impact of promoting researchers working in more feminized topics, we provide causal evidence on the role of research alignment. We find that topic alignment does increase the number of female PhD students, but gender has an additional and substantially larger effect, pointing to mechanisms specific to gender — such as mentoring, role-model effects, and access to professional networks — operating alongside research affinity.

Fourth, we contribute to the broader literature on gender and organizations. Much of this work examines how female representation in leadership affects outcomes at a single career stage, such as hiring, promotion, or retention (e.g., Kunze and Miller, 2017; Azmat and Boring, 2020; Folke and Rickne, 2022; Cullen and Perez-Truglia, 2023). Our analysis allows us to trace how representation effects propagate across the academic pipeline, from faculty composition to PhD training to graduates' subsequent careers, over a horizon of fifteen years. This long-run perspective complements firm-level studies, which typically observe neither such downstream dynamics nor effects over such a long period.

The remainder of the article is organized as follows. Sections 2 and 3 describe the institutional setting and our linked dataset on faculty careers, departmental composition, and PhD outcomes. Section 4 presents descriptive evidence motivating the causal analysis, while Section 5 outlines our identification strategy. Section 6 presents the main results, and Section 7 examines the mechanisms behind our findings and discusses alternative interpretations. Section 8 concludes.

2 Institutional Context

2.1 Organization of Spanish Academia

Our analysis focuses on Spain's public university system over the period 1985–2023. Spain has 48 public universities, which account for over 90% of all faculty members and students during this period.[5] In 2010, these universities employed approximately 100,000 academic staff, of whom about 9% were full professors, 28% associate professors, 17% tenured assistant

5. We do not consider two additional public institutions, Universidad Internacional Menéndez Pelayo (UIMP) and the Universidad Internacional de Andalucía (UNIA), which do not employ their own permanent academic staff, relying instead on faculty seconded from other universities.

professors (non-civil servants), and 12% non-tenured assistant professors.[6] In addition, the remaining 35% held non-research positions including part-time teaching fellows and teaching-track staff not requiring a doctoral degree.[7] Faculty members are organized within 188 nationally defined fields (*áreas de conocimiento*) and, within universities, these fields typically correspond to academic departments, though the correspondence is not always exact. In some universities, a single department spans multiple fields (e.g., management, accounting, and marketing), but departments comprising several fields generally organize hiring and promotion at the field level. On the other hand, in some large universities there might be several departments within the same field (e.g., multiple economics departments in Universidad Complutense).

2.2 The *Habilitación* System

Prior to the early 2000s, Spanish universities enjoyed substantial autonomy in hiring and promotion, characterized by a high degree of academic inbreeding among both PhD students and faculty. According to a 2006 survey, approximately two-thirds of PhD students had completed their undergraduate degree at the same institution, and nearly three-quarters of faculty were employed at the university where they earned their PhD (Cruz Castro et al., 2006). In response, in 2002 the Spanish government introduced a two-stage promotion system aimed at increasing transparency and meritocracy. Candidates for Associate and Full Professor positions were first required to obtain a national-level qualification through the *habilitación* exam, after which they could apply for positions at individual universities. This system resembles centralized qualification procedures used in other countries, such as France's *agrégation* and Italy's *abilitazione scientifica nazionale*.

The *habilitación* system was in place from 2002 to 2008 and had two features worth noting for our identification strategy. First, the number of national qualifications was capped based

6. Throughout the paper, Full Professor refers to *Catedrático de Universidad* and Associate Professor to *Profesor Titular de Universidad*. Other permanent positions include *Profesor Contratado Doctor* (tenured assistant professor) and *Profesor Ayudante Doctor* (non-tenured assistant professor).

7. Source: *Estadística de la Enseñanza Universitaria en España. Curso 2010/2011*, Instituto Nacional de Estadística.

on estimates of available positions across Spanish universities. Due to the limited number of nationally qualified individuals, most of those who passed the exam were subsequently promoted by their home universities with little external competition. Second, evaluation committees were formed through a random draw from the pool of eligible professors in the corresponding field, with no rules precluding conflicts of interest between evaluators and candidates.

The promotion process began with a call for applications issued by the Ministry, to which applicants had twenty days to respond. After the deadline, committee members were selected by random draw from the pool of eligible evaluators in the relevant field and rank, drawn from public universities and public research institutes (see details in Appendix B). Eligibility required meeting minimum research productivity thresholds set by the national research evaluation agency. Each evaluation committee consisted of seven members: committees for Associate Professor evaluations included three Full Professors and four Associate Professors, while committees for Full Professor evaluations were composed entirely of Full Professors. A reserve committee of equal size was also drawn. Participation was mandatory, and fewer than 2% of selected evaluators were replaced.[8]

Candidates were evaluated through a curriculum vitae presentation and a research seminar; Associate Professor candidates also completed a teaching demonstration. Decisions were made by majority vote. The duration of the evaluation process varied across fields and cohorts: roughly one-third of committees concluded within one year, most within two years.

Following the final *habilitación* rounds (2006–2008), Spain transitioned to the current *acreditación* system. While the new system retains the two-stage structure, it removed the cap on the number of candidates who may qualify at the national level, and in practice roughly half of applicants qualify, compared to only about 10–12% under *habilitación*. The committee selection process was also reformed, precluding the identification strategy we use for the

8. Withdrawals were permitted only if they were serving in a Spanish public administration role or in cases of close personal ties to candidates (Ley de Procedimiento Administrativo 30/1992, article 28, available at https://www.boe.es/eli/es/l/1992/11/26/30/con, last accessed 11 July 2025). Based on our own calculations, resignations owing to connections with candidates were exceedingly rare. For instance, of the 832 professors assigned to evaluate their own Ph.D. students, only 22 opted out, a proportion similar to the overall withdrawal rate.

habilitación period, as committee members are now selected to avoid conflicts of interest. Promotion activity experienced two major slowdowns during our study period: first between 2004 and 2006, as the *habilitación* system created a bottleneck in qualifications, and again between 2013 and 2016, when the government imposed a nationwide hiring freeze on public universities as part of austerity measures.[9]

2.3 Doctoral Training

Doctoral education in Spain is organized within a nationally regulated, program-based framework.[10] Doctoral candidates are admitted to officially recognized doctoral programs housed within universities and coordinated through Doctoral Schools, rather than being recruited directly by individual faculty members.

Each doctoral program is overseen by an Academic Committee that determines admissions, assigns supervisors, approves research plans, and monitors progress through periodic evaluations. A central feature of the Spanish system is that lead supervisors must meet accreditation and research-activity requirements defined by universities and national quality-assurance guidelines. While junior faculty, including non-tenured assistant professors, may participate in supervision, primary supervisory responsibility is typically held by tenured senior faculty.

Doctoral programs vary substantially across disciplines in size, structure, and funding. In the natural sciences, engineering, and medicine, many doctoral students are funded through competitive contracts or research grants tied to specific projects and research groups, so that recruitment partly depends on grant funding. In the humanities and parts of the social sciences, doctoral cohorts tend to be smaller, funding is more heterogeneous, and part-time enrollment more common. Medicine and health sciences represent the upper end of the size distribution, reflecting the scale of biomedical research and the integration of universities with hospitals and public research institutes. These differences in program structure may shape

9. See El País (2013) and The Guardian (2013).

10. During the period we study, PhD training was governed by successive national decrees, culminating in the comprehensive reform introduced by Royal Decree 99/2011, which standardized doctoral programs across universities and disciplines, available at https://www.boe.es/eli/es/rd/2011/01/28/99/con, last accessed 7 January 2026.

how female faculty presence affects PhD outcomes, a possibility we explore by examining heterogeneity across broad field groups.

3 Data

We construct a linked dataset combining faculty careers, departmental composition, and PhD outcomes. Building on data from Bagues et al., 2017 on candidates, evaluators, committee assignments, and evaluation outcomes under the *habilitación* system, we augment it with administrative information on promotions from the *Boletín Oficial del Estado* (BOE), multiple bibliometric databases (OpenAlex, Scopus, Dialnet), and the near-universe of registered PhD dissertations from TESEO.

The resulting dataset allows us to follow academic careers and doctoral training at the department level, covering all 48 public universities and 188 research fields (*áreas de conocimiento*), with fields harmonized to account for changes in official definitions over time. Appendix H provides additional details on data sources and harmonization procedures.

The section proceeds in three steps: (i) we describe the structure of the academic pipeline in Spanish universities, (ii) we characterize the institutional setup of the *habilitación* system, and (iii) we provide benchmarks for interpreting both observational correlations and causal estimates in the sections that follow.

Departments, faculty composition, and promotions. We define departments as the set of researchers in a given research field within a given university. As discussed in Section 2, these units closely correspond to actual departments or, in the case of multifield departments, to the sub-units responsible for hiring and promotion within a given field. Using individual career trajectories, we assign each researcher to the department in which they first obtain a tenured position and assume that this affiliation remains fixed until age 65, unless the individual is subsequently hired by another public university. This assumption reflects the high degree of employment stability in the Spanish academic system and allows us to construct consistent department-level faculty stocks over time.

To assess the accuracy of the resulting department-level database, we validate our measures using official statistics from the Spanish Ministry of Education, which report departmental size and gender composition for the period 2011–2018 for departments with at least five female or male faculty members. The comparison shows a correlation above 90%, suggesting that measurement error is limited.

Our dataset covers 5,422 departments over the period 2000–2023, of which 4,000 had at least one faculty member participate in a qualification exam and form the basis of our causal analysis. Appendix Figure A1 shows the distribution of tenured professors by department size in 2000. The median professor works in a department with 11 tenured faculty members, of whom roughly 20% are full professors. However, department size is highly skewed. Notably, approximately 1,000 researchers are the only tenured faculty member in their field–university unit, while at the upper end of the size distribution, only 20 departments have more than 50 tenured professors.

Gender composition varies across fields. The median professor in Science, Technology, Engineering, Mathematics, and Medicine (STEM+M) is in a department where 25% of tenured faculty are women, compared to 30% in Social Sciences and Humanities (SSH) (see Appendix Figure A2). Female representation also differs markedly by academic rank, and has increased over time at both levels. Among Associate Professors, the share of women rose from 32% in 2000 to 43% in 2023, while among Full Professors it increased from 13% to 28% over the same period (Appendix Figure A3).

Appendix Figure A4 documents substantial time variation in promotion flows, reflecting both the *habilitación* bottleneck and the 2013–2016 hiring freeze mentioned in Section 2. In contrast, department size is relatively stable over time. The mean number of Associate Professors increases only modestly, from five in 2000 to six in 2023, while the mean number of Full Professors remains stable at about two throughout the period.

PhD students and early-career outcomes. We assemble a dataset covering 203,542 doctoral theses defended in Spanish public universities between 1990 and 2017, after excluding approximately 8% of theses with missing information on research field, graduation year, or advisor. As shown in Panel A of Appendix Table A1, 47% of PhD graduates during this period were women, while 21% of PhD students had a female supervisor. Women are significantly more likely to be advised by female supervisors than men (27% versus 16%).

The average time to complete the degree is approximately five years.[11] Around 50% of PhD graduates remain academically active—defined as producing at least one research output five years after graduation—consistent with survey evidence indicating that academic careers absorb less than half of PhD graduates in Spain, with many finding employment in the private sector, public research organizations, healthcare, and government (Auriol et al., 2013). Around 10% of graduates are promoted to Associate Professor within ten years of completing their PhD.

Gender gaps in these outcomes are substantial. Compared to men, female PhD graduates take slightly longer to complete their degrees, are less likely to remain academically active, and are less likely to be promoted to Associate Professor within ten years. These gaps are particularly pronounced in STEM+M fields, where women are 7% less likely than men to remain academically active and 45% less likely to be promoted to Associate Professor within ten years of graduation (Panel B of Appendix Table A1). In the Social Sciences and Humanities, gender differences are more muted: women are slightly more likely to remain academically active and slightly less likely to be promoted to Associate Professor than men (Panel C of Appendix Table A1).

We also document substantial gender differences in research productivity during and after the PhD. Women tend to produce fewer research outputs both during and after their doctorate, but with no difference in the average Article Influence Score (AIS), suggesting differences in publication quantity rather than quality.[12]

11. We directly observe enrollment year for around 30% of graduates. For the remainder, we impute enrollment year based on department, cohort, and gender.

12. AIS is a journal-level metric similar to the impact factor but adjusted for the influence of citing journals and the number of references, excluding self-citations. It is normalized to an average of one across Web of Science journals.

Gender differences in research specialization. To distinguish the role of gender from research specialization, we characterize the gender composition of research topics and examine the extent to which women and men concentrate in different research areas. We construct separate measures for faculty and PhD students, using different data sources for each.

For faculty and exam candidates, we use OpenAlex to identify approximately 4,500 research topics. For each topic-year, we compute the share of female authors publishing in that topic over the preceding five years, excluding the current year.[13] We aggregate these topic-level indices to the paper and author level. An author's yearly index reflects the average gender composition of the research topics in which they publish, capturing the gender structure of their research areas rather than their own gender. To give a sense of magnitude, in 2000 the topic *Early Childhood Education and Development* had a feminization index of 0.60, while *Evolutionary Game Theory and Cooperation* had an index of 0.22.

For PhD students, we rely instead on the UNESCO Thesaurus classification, which all graduates in Spain must report at the time of dissertation registration. The taxonomy includes 24 broad two-digit fields (e.g., 24: *Life Sciences*), 250 four-digit categories (e.g., 2409: *Genetics*), and 2,153 six-digit subfields (e.g., 240901: *Embryology*).

Women and men specialize in different research areas on average, as shown in Appendix Figure A5, which plots the distributions of the topic gender composition index for professors, exam candidates, PhD graduates, and their advisors. While women's distributions are shifted toward more feminized topics, the overlap is substantial.

We analyze the evolution of gender segregation in research topics over time using PhD graduates data, where the UNESCO classification provides consistent measurement across cohorts. We compute the Duncan segregation index, which measures the fraction of graduates who would need to reallocate across research topics for the gender distributions to be identical.[14] Appendix Figure A6 plots the segregation index over time at three levels of

13. By excluding the current year, the feminization index used to characterize a given publication is not mechanically influenced by the gender of its own authors. For publications not directly indexed in OpenAlex, we map concept labels from the Microsoft Academic Graph (MAG) classifier to the OpenAlex taxonomy. See Appendix Section H.5 for more details.

14. Formally, $D = \frac{1}{2}\sum_{j=1}^{J}\left|\frac{f_j}{F} - \frac{m_j}{M}\right|$, where f_j and m_j denote the number of female and male graduates in

aggregation. Gender segregation is sizable and persistent at all levels, with larger values at finer classifications. At the six-digit level, the index declines from roughly 0.40 in the 1990s to just above 0.30 by 2015, implying that about one-third of graduates would need to change topics to match the distribution of the opposite gender. While segregation has declined meaningfully at the six-digit level, this does not reflect a shift in broad research orientation — at the four-digit and two-digit levels, segregation remains stable at around 0.20 and 0.15–0.20 respectively, suggesting that the narrowing of fine-grained segregation reflects reshuffling within broad fields rather than a convergence in overall research orientation.

Candidates in the *habilitación* system. Our dataset covers 967 *habilitación* exams across 174 research fields.[15] Roughly half correspond to Associate Professor positions (465) and half to Full Professor positions (502). In total, we observe 31,243 applications. Candidates submit two applications on average, reflecting both repeated attempts across exam waves and simultaneous applications to closely related fields.[16]

The qualification process is highly selective at both ranks. As shown in Appendix Table A2, only about 11% of Full Professor candidates and 12% of Associate Professor candidates qualify. Applicants to Full Professor positions are substantially more senior, with 43 publications on average (journal articles, books, or book chapters) versus 21 for Associate Professor applicants. Average journal impact, measured by the Article Influence Score, is similar across ranks (0.83 for Full Professor candidates and 0.87 for Associate Professor candidates).

Professional connections to evaluators are common. Approximately one-third of applicants have at least one *strong tie* with a committee member, defined as prior doctoral supervision, co-authorship, or shared departmental affiliation. Such ties are more prevalent in Full Professor exams (35%) than in Associate Professor exams (29%), largely reflecting shared departmental affiliation (32% versus 27%). Co-authorship ties are present in 7% of Full

topic j, and F and M are the respective totals.

15. We retain the official field classification in place at the time of the *habilitación* exams and map these fields to the harmonized 188-field classification used elsewhere in the paper. We exclude 11 fields in which the pool of eligible evaluators is too small to guarantee random committee assignment.

16. In the analysis, we exclude 42 applications from eight candidates who reapplied after being promoted in a previous wave and were subsequently removed from the selection process.

Professor and 4% of Associate Professor applications, while advisor–advisee links occur in 3% and 2% respectively.

Gender differences in qualification outcomes are present at both ranks. Among Full Professor candidates, women qualify at a rate of 9%, compared to 11% for men; among Associate Professor candidates, the corresponding rates are 11% and 12%. These gaps are accompanied by differences in observable research output: on average, female candidates have fewer publications than male candidates and lower average Article Influence Score. By contrast, women and men are similarly exposed to evaluators with whom they have strong ties – although the intensity of these ties is slightly lower for women – suggesting that women are not systematically disadvantaged in their exposure to connected evaluators.

Appendix Figure A7 documents the link between qualification outcomes and subsequent promotions. Among successful candidates, more than 90% are promoted within two years of the exam, and among those promoted, 86% obtain their position at the same department where they were based at the time of qualification, reflecting the near-mechanical link between qualification and promotion during the *habilitación* period. Among unsuccessful candidates, a substantial share eventually reapplies and qualifies in later rounds, leading to a gradual catch-up in promotion rates over time. This catch-up is more pronounced for Full Professor candidates, consistent with the fact that unsuccessful Full Professor candidates retain their Associate Professor tenure and face fewer barriers to reapplication.

Departments in the *habilitación* system. The *habilitación* system generates substantial variation in the gender composition of departments, which we exploit to study downstream effects on faculty composition and PhD outcomes. Approximately 75% of university departments in Spain submit at least one application during our study period, and these constitute our analysis sample (Appendix Table A3). Participating departments tend to be larger than average, employing 11–13 tenured faculty compared to an average of approximately 6 among all departments nationally, reflecting that larger departments are more likely to have candidates seeking promotion (Panel A of Appendix Table A4). Female candidacies are considerably less common than male candidacies, particularly at the Full Professor level:

only 29% of departments submit at least one female Full Professor candidate versus 58% submitting a male candidate. At the Associate Professor level, the gap is smaller (52% versus 64%). Conditional on participating in a given exam, departments typically submit one to two candidates.

Reflecting the competitiveness of the qualification process, successful qualifications are rare and predominantly male. Appendix Table A3 shows that only 7% of departments qualify a female Full Professor candidate versus 20% for males; at the Associate Professor level, the corresponding figures are 14% and 22%.

Panel B of Appendix Table A4 reports post-exam outcomes. Over the ten years following a Full Professor exam, participating departments promote on average 1.7 men and 0.7 women to Full Professor. Following Associate Professor exams, departments promote 2.4 men and 1.5 women to Associate Professor. Both types of departments graduate approximately 13–15 PhD students of each gender, of whom roughly half remain academically active.

4 Descriptive Evidence on Gender and the Academic Pipeline

Before turning to the causal analysis, we document descriptive associations suggestive of trickle-down dynamics. These patterns help motivate the empirical strategy and the mechanisms we explore in the causal analysis.

We focus on two descriptive patterns. First, we examine whether female representation at lower career stages is related to the presence of women higher up the academic ladder and to the gender composition of the research areas in which faculty specialize. Second, we study whether female PhD students are more likely than their male peers in the same department and cohort to be supervised by women, and examine how much of this pattern reflects gender differences in research specialization.

Share of women at different levels of the academic ladder. We first examine whether the shares of female Full Professors and Associate Professors in a department, as well as the feminization of their research topics, predict the share of women among new PhD students and newly promoted Associate Professors in the following year. Table 1 reports OLS estimates using department–year data, controlling for field and year fixed effects. The sample includes only departments with at least one Associate Professor and one Full Professor in a given year, ensuring that the gender composition measures are defined. Columns 1–2 focus on PhD enrollment and columns 3–4 on newly promoted Associate Professors.

There is a significant positive association between the presence of senior female faculty and women's entry at earlier stages of the pipeline (columns 1 and 3). A 10–percentage-point increase in the share of female Associate Professors is associated with a 0.4–percentage-point increase in the female share of new PhD students and a 0.8–percentage-point increase among newly promoted Associate Professors; the corresponding figures for female Full Professors are 0.3 and 0.5 percentage points. Relative to baseline female shares of 48% among new PhD students and 37% among promoted Associate Professors, these are modest annual shifts that can nonetheless compound into meaningful changes over time given slow faculty turnover.

To assess whether these correlations reflect gender segregation across research fields, columns 2 and 4 control for the feminization of senior faculty research areas, measured as the average gender composition of the research topics they work on. The feminization of Full Professors' research is not significantly associated with the share of women among PhD graduates, nor with the share among newly promoted Associate Professors. In contrast, the feminization of Associate Professors' research correlates positively with both outcomes, though the effects are modest in magnitude. Importantly, adding these controls leaves the coefficients on senior female representation essentially unchanged, suggesting that field-level gender segregation explains only a limited share of the observed correlation between the share of women at different levels.

Gender Homophily in PhD Supervision and Research Specialization. Second, we examine whether female PhD students, relative to male students in the same department and cohort, are more likely to be advised by female professors or by professors whose research lies in more feminized areas. Table 2 reports OLS estimates from regressions of a female graduate indicator on the advisor's gender, controlling for year and department fixed effects (column 1). Female advisors are substantially more likely to supervise female students: the share of female PhD graduates is 9.9 percentage points higher among those with a female advisor, relative to a baseline of 43%.

Female PhD students are also more likely to be advised by professors working in more feminized research areas. Column 2 of Table 2 shows that a one-standard-deviation increase in an advisor's topic feminization index is associated with a six-percentage-point higher probability that the supervised student is female. Columns 3 and 4 show that both coefficients remain large and statistically significant when included simultaneously and when department–year fixed effects are added, reinforcing the conclusion that both advisor gender and research interests contribute to advisor-student matching.

5 Empirical Strategy

This section presents the instrumental-variables (IV) strategy that we use to identify the causal effect of female qualifications on individual careers and departmental outcomes. The approach builds on Zinovyeva and Bagues, 2015, who show that candidates randomly assigned to committees that include their academic connections are significantly more likely to qualify. The realized presence of a connection in the committee thus generates variation in qualification outcomes that is as good as random, conditional on the expected probability of such connections.

Our identification strategy compares the number of connections a candidate has in the evaluation committee with the connections they were expected to have based on how many connections they had in the pool of eligible evaluators. The deviation between the two captures variation arising from the random draw of committee members. This recentering approach, which Borusyak and Hull, 2023 formalize in a general setting, ensures that we isolate exogenous shocks in evaluator composition while controlling for any nonrandom exposure to these shocks. Conditioning on expected connectedness in this way eliminates potential bias arising from systematic differences in candidate networks or departmental structures.[17]

We first use this individual-level design to estimate how exogenous qualification affects subsequent promotion prospects, both validating the research design and providing estimates of the causal effect of qualification on individual careers. We then aggregate this variation to the department level to study how exogenous promotions of women and men propagate within departments, affecting faculty composition, PhD training, and the retention of female researchers.

Two features of the institutional environment strengthen our research design. First, during the *habilitación* period, binding constraints on the number of qualifications meant that qualification translated almost directly into promotion — with more than 90% of successful candidates promoted within two years — generating a strong first stage. Second, the hiring freeze between 2013 and 2016 aids identification of lasting effects by limiting the scope for catch-up among unsuccessful candidates, making the treatment more persistent over time, though by suppressing promotion activity more broadly it may also attenuate medium-run effects on faculty composition while leaving long-run effects and PhD outcomes unaffected. The following sections describe the identifying assumptions, specification choices, and interpretation of the key coefficients.

5.1 Individual-level identification

We begin by assessing whether candidates who qualify due to a favorable draw of committee members are subsequently more likely to be promoted. In the short run, this is largely mechanical, as noted above, given that qualification almost directly translated into promotion during the *habilitación* period. The more informative horizon is the long run, where the effect measures the extent to which the initial advantage persists despite the fact that unsuccessful candidates could, and frequently did, reapply and qualify in later rounds.

17. In a related paper, Bagues et al., 2017 exploit variation in the gender composition of committees, but find that the effect of committee gender composition on qualification outcomes is too small to generate sufficient first-stage variation for our purposes.

First stage. Let $Q_{i,e,t}$ denote an indicator for whether candidate i qualifies in exam call e at time t, and $C_{i,e}$ the measure of connectedness between candidate i and their committee.[18] Here, t refers to the year in which qualification results are announced, typically one to two years after the exam was initially called. Connectedness is measured as the total number of connections a candidate has with each evaluator, summed across all committee members, where connections include coauthorship, institutional affiliation, and doctoral supervision. Each connection type counts separately: if a candidate is both affiliated with and has coauthored with a given evaluator, this contributes two connections. The resulting measure $C_{i,e}$ thus captures both the number and intensity of connections. Appendix D details the construction of connectedness and reports robustness checks based on alternative definitions.

Given the random assignment of evaluators described in Section 2, it is possible to compute each candidate's expected connectedness $\mathbb{E}[C_{i,e}]$ under random assignment (see Appendix B for details). The deviation of the realized committee composition from its expectation,

$$Z_{i,e} \equiv C_{i,e} - \mathbb{E}[C_{i,e}],$$

has expected value zero and reflects only the random draw of committee members (Borusyak and Hull, 2023).

The first-stage equation is:

$$Q_{i,e,t} = \alpha_0 + \alpha_1 Z_{i,e} + \alpha_2 \mathbb{E}[C_{i,e}] + \alpha_3 \mathbf{X}_{i,e} + \mu_e + \lambda_t + \varepsilon_{i,e,t}, \tag{1}$$

where, to increase precision, $\mathbf{X}_{i,e}$ includes predetermined characteristics of the candidate (such as research productivity at the time of the exam), μ_e are exam fixed effects, and λ_t are fixed effects for the year in which qualification results are released. The coefficient α_1 captures the causal effect of an additional connection to a committee member on the probability of qualification.

18. Throughout the paper, we use the composition of the initially drawn committee as the treatment, before any resignations took place. As explained in Section 2, resignations were very rare.

Second stage. To estimate the effect of qualification on promotion to the corresponding rank (Associate or Full Professor) within k years, we estimate:

$$P_{i,e,t+k} = \beta_0 + \beta_1 \widehat{Q}_{i,e,t} + \beta_2 \mathbb{E}[C_{i,e}] + \beta_3 \mathbf{X}_{i,e} + \mu_e + \lambda_t + \eta_{i,e,t+k}, \tag{2}$$

where $\widehat{Q}_{i,e,t}$ is the fitted value from the first-stage regression, with $Q_{i,e,t}$ instrumented by $Z_{i,e}$ — the deviation between realized and expected connectedness. We estimate Equation (2) by 2SLS, both on the overall sample and separately by gender. Standard errors are clustered at the exam level.

Identifying assumptions and estimand's interpretation. The validity of our instrumental variables approach rests on three standard conditions. The first is that the variation in committee composition we exploit is unrelated to candidate characteristics, a condition that holds by design given the random assignment of evaluators and is confirmed empirically in Appendix Table C1.

Beyond random assignment, we also require that committee composition affects promotions only through qualification outcomes. The most plausible violation would be if connected evaluators, randomly assigned at the national level, directly influenced subsequent hiring decisions at individual universities. This channel is unlikely for two reasons. First, national evaluations were one-off events with no formal role in university-level hiring. Second, the promotion process was largely mechanical: universities faced a tightly constrained number of positions and promoted more than 90% of successful candidates within two years of qualification, leaving little room for evaluators to exert additional influence over outcomes.

A final concern — corresponding to the monotonicity assumption — is whether connections ever hurt rather than help candidates, for instance due to strained professional relationships. While such cases are conceivable, actively disadvantaging a connected candidate would run counter to prevailing academic norms and risk reputational costs, making systematic negative effects unlikely. The first-stage estimates also provide indirect reassurance: a non-negligible share of defiers would attenuate the estimated relationship between committee composition and qualification, which is difficult to reconcile with the size and precision of our first-stage results.

Under these conditions, the coefficient β_1 in Equation (2) identifies the Local Average Treatment Effect (LATE) of qualification on promotion for candidates whose qualification status is shifted by random variation in committee composition. In the Spanish *habilitación* system, where only about 10–12% of candidates qualify and roughly one-third have at least one connection to the committee, compliers are candidates who fall just below the qualification threshold but are pushed above it by a favorable committee draw. One would therefore expect compliers to be of high quality — a prediction confirmed by the evidence in Appendix E. Following Imbens and Rubin, 1997, we characterize the complier population by estimating the shares and observable characteristics of always-takers, compliers, and never-takers. Compliers account for approximately 8.5% of candidates, with always-takers and never-takers accounting for 9% and 82.5% respectively — reflecting the highly competitive nature of the *habilitación*. In terms of number of publications, always-takers are around 0.45 standard deviations more productive than never-takers and compliers are around 0.14 standard deviations more productive (Appendix Table E2 and Appendix Figure E2), confirming that our estimates capture the effect of qualification for highly capable candidates who are close to the margin, rather than for the average applicant.

This estimand is particularly policy-relevant. Our estimates do not capture the effect of promoting candidates who would have succeeded regardless, nor the effect of promoting candidates far below the qualification threshold. Instead, they identify the impact of resolving qualification uncertainty for candidates at the margin — those strong enough to be competitive but whose outcomes remain sensitive to the composition of their evaluation committee. These are precisely the candidates whose career progression is likely to be most responsive to marginal institutional interventions aimed at reducing promotion frictions, making our estimates directly relevant for the design of gender equity policies in academic hiring.

Two further considerations bear on the interpretation of the LATE. First, with a continuous instrument, treatment effect heterogeneity may affect interpretation (Mogstad and Torgovitsky, 2024). We address this concern by showing that our results are robust to alternative instrument definitions, including a binary specification reported in Appendix D.

Second, connected evaluators may behave strategically, exerting influence more strongly when qualification is expected to translate directly into a promotion at the candidate's home department. This does not threaten identification, but it does clarify what the LATE captures: candidates whose qualification is both marginal and promotion-relevant. Importantly, the institutional features of the *habilitación* system limit the practical scope of this concern. Because the number of qualifications was centrally determined, qualification itself was the primary competitive bottleneck, giving departmental representatives strong incentives to support candidates regardless of immediate promotion capacity. Systematic strategic withholding of support was therefore unlikely.

5.2 Department-level identification

Next we examine how individual qualification outcomes propagate to the department level. The department-level analysis builds directly on the individual design by aggregating the same random variation in committee composition. In the short run, the department-level effects are largely mechanical, reflecting the direct promotion of the exogenously qualified individual. Over time, however, the estimated effects capture two compounding forces: the net individual effect — accounting for the gradual catch-up of initially unsuccessful candidates — and any knock-on effects generated by the initial promotion, such as changes in subsequent hiring, PhD training, and the retention of junior researchers.

We proceed in two steps. First, we study the effect of a qualification shock on department size and promotions, independently of the gender of who qualified, establishing the aggregate relevance of individual qualification events. We then decompose these shocks by gender to identify how female and male qualifications differentially shape departmental dynamics.

Impact of qualification shocks, unconditional on gender. Our estimation sample consists of all department-exam pairs (j, e) in which at least one candidate from department j applied to exam e. Let $N_{j,e}$ denote the total number of applicants from department j in exam e, and $Q_{j,e,t}$ the number who qualify at time t. Let $Y_{j,e,t+k}$ denote a departmental outcome measured over the period $[t, t+k]$, such as the cumulative number of promotions or PhD graduates during this period. Throughout, both treatments and outcomes are defined in levels rather than shares, as departments typically have only one or two candidates per exam.

Following the individual-level analysis, we construct a department-level instrument by averaging the individual-level deviations of realized from expected connectedness:

$$\overline{Z}_{j,e} \;=\; \frac{1}{N_{j,e}} \sum_{i=1}^{N_{j,e}} \Big(C_{i,e} - \mathbb{E}[C_{i,e}] \Big),$$

We also define the average expected connectedness, which enters as a control in the first-stage regression to absorb systematic differences in network density across department-exam cells:

$$\overline{\mathbb{E}[C]}_{j,e} \;=\; \frac{1}{N_{j,e}} \sum_{i=1}^{N_{j,e}} \mathbb{E}[C_{i,e}].$$

Averaging individual deviations rather than summing them addresses the fact that a single connected evaluator can only favor a limited number of applicants, even when shared by multiple candidates from the same department. In the extreme case in which all candidates share one connected evaluator, the favorable draw effectively generates one favorable outcome rather than one per candidate.

The first-stage equation is:

$$Q_{j,e,t} = \gamma_0 + \gamma_1 \overline{Z}_{j,e,t} + \gamma_2 \overline{\mathbb{E}[C]}_{j,e,t} + \theta' W_{j,e,t} + \mu_e + \lambda_t + v_{j,e,t}, \tag{3}$$

and the corresponding second stage is:

$$Y_{j,e,t+k} = \delta_0 + \delta_1 \widehat{Q}_{j,e,t} + \delta_2 \overline{\mathbb{E}[C]}_{j,e,t} + \theta' W_{j,e,t} + \mu_e + \lambda_t + u_{j,e,t+k}, \tag{4}$$

where $Q_{j,e,t}$ is instrumented by $\overline{Z}_{j,e,t}$, so that δ_1 captures the causal effect of one additional exogenous qualification on departmental outcomes. The vector of controls $W_{j,e,t}$ includes the number of applicants, pre-exam department faculty size, pre-exam PhD cohort size, and promotion rates in the years preceding the exam. While these controls are not required for identification — since the instrument is defined as a deviation from expected connectedness — they absorb variation in departmental characteristics that may correlate with outcomes and improve precision. Similarly, exam and qualification-year fixed effects (μ_e, λ_t) absorb variation across exams and period-specific shocks. The model is estimated by 2SLS, with standard errors clustered at the department level.

The validity of the department-level IV strategy rests on the same conditions as the individual-level analysis, which we discuss briefly in turn. Exogeneity follows directly from the random assignment of evaluators: the average deviation in committee composition $\overline{Z}_{je}$ is by construction unrelated to predetermined departmental characteristics, a claim supported by the balance evidence in Appendix Table C2 and Appendix Table C3. The exclusion restriction requires that random variation in committee composition affects departmental outcomes only through its impact on qualification decisions. The arguments made at the individual level apply here as well: national evaluations were one-off events with no formal role in university-level hiring, and the promotion process was largely mechanical. A potential additional concern at the department level is that a lucky draw could simultaneously qualify multiple candidates from the same department, generating internal competition if positions are scarce. However, given that most departments sent only one or two candidates per exam, and that the number of available positions was centrally determined, such conflicts were rare and unlikely to systematically affect departmental outcomes. Monotonicity requires that a favorable committee draw never reduces the probability of qualification for any candidate in the department, an assumption supported by the same arguments made at the individual level.

Impact of qualification shocks, by gender. We then extend this framework to separately identify the effects of female and male qualifications on departmental outcomes. Let $Q^f_{j,e,t}$ and $Q^m_{j,e,t}$ denote the numbers of qualified women and men from department j in exam e at time t, and $N^f_{j,e}$ and $N^m_{j,e}$ the corresponding numbers of female and male applicants.

We construct gender-specific instruments by averaging individual deviations of realized from expected connectedness within each gender:

$$\overline{Z}^g_{j,e} \;=\; \frac{1}{N^g_{j,e}} \sum_{i=1}^{N^g_{j,e}} \Big(C_{i,e} - \mathbb{E}[C_{i,e}] \Big), \qquad g \in \{f, m\},$$

with $\overline{Z}^g_{j,e} = 0$ if $N^g_{j,e} = 0$, and gender-specific controls $\overline{\mathbb{E}[C]}^g_{j,e}$ defined analogously and included in all specifications.

The first stage jointly instruments the number of qualified women and men, using both gender-specific instruments simultaneously:

$$\begin{aligned} Q^g_{j,e,t} =& \gamma^g_0 + \gamma^g_1 \overline{Z}^f_{j,e,t} + \gamma^g_2 \overline{Z}^m_{j,e,t} + \gamma^g_3 \overline{\mathbb{E}[C]}^f_{j,e,t} + \gamma^g_4 \overline{\mathbb{E}[C]}^m_{j,e,t} \\ & + \theta^{g\prime} W_{j,e,t} + \mu_e + \lambda_t + v^g_{j,e,t}, \quad g \in \{f, m\}. \end{aligned} \tag{5}$$

The second stage for a generic departmental outcome — capturing either the gender composition of faculty or the career outcomes of PhD graduates — is:

$$Y_{j,e,t+k} = \delta_0 + \delta^f_1 \widehat{Q}^f_{j,e,t} + \delta^m_1 \widehat{Q}^m_{j,e,t} + \delta^f_2 \overline{\mathbb{E}[C]}^f_{j,e,t} + \delta^m_2 \overline{\mathbb{E}[C]}^m_{j,e,t} + \theta' W_{j,e,t} + \mu_e + \lambda_t + u_{j,e,t+k}. \tag{6}$$

As in the individual-level analysis, δ^f_1 and δ^m_1 identify department-level Local Average Treatment Effects — the causal effects on departmental outcomes in departments where random committee composition shifts the number of qualified female and male candidates, respectively.

6 Main Results

We organize the causal evidence in four parts. Section 6.1 examines the impact of qualifications on individual and departmental promotions. Section 6.2 studies the effects on PhD graduate outcomes. Section 6.3 explores heterogeneity by baseline female representation and field group. Section 6.4 examines the role of research orientation as a complementary channel.

6.1 Impact of Qualifications on Promotions

We begin by showing that exogenous qualification has persistent effects on individual academic careers, particularly at the Associate Professor level. We then examine whether these individual-level shocks propagate into lasting changes in departmental gender composition.

6.1.1 Individual-level Evidence

The core idea behind our individual-level analysis is simple: candidates who happen to be assigned to a committee that includes their academic connections are more likely to qualify, and this lucky draw translates into lasting career advantages.

Appendix Table C4 reports first-stage estimates, confirming that committee composition matters strongly for qualification outcomes, consistent with the earlier evidence in Zinovyeva and Bagues, 2015. Throughout, we report effects in percentage points (p.p.) and standard errors (s.e.) in parentheses. In Full Professor exams, each additional strong connection among the seven evaluators raises the probability of qualification by 5.2 p.p. (s.e. = 0.5), representing a 47% increase relative to the baseline qualification rate. In Associate Professor exams, the effect is somewhat larger at 6.4 p.p. (s.e. = 0.6), or 59%. These estimates are highly significant and virtually unchanged when controlling for prior research productivity, as expected under random committee assignment.

Exploiting this first-stage variation, Figure 1 plots 2SLS estimates of the effect of an exogenous qualification on promotion probabilities over a 15-year horizon, separately by rank. In the short run, the pattern is similar across ranks: one year after the committee decision, exogenously qualified candidates are approximately 75 p.p. (s.e. = 7) more likely to have been promoted. This falls somewhat short of the 90% promotion rate within one year documented in Appendix Figure A7, reflecting some early catch-up by initially unsuccessful candidates. The key difference across ranks is the persistence of the shock to career progression.

For Associate Professor candidates, the career advantage is large and remarkably durable. Even ten years after the exam, candidates who qualified due to a favorable committee draw remain 54 p.p. (s.e. = 11) more likely to have been promoted than otherwise similar candidates who did not, with the effect settling at 40 p.p. (s.e. = 11) after fifteen years. The top panels of Figure 2 show that this persistence is more pronounced for women than for men. Ten years after the qualification decision, the difference in promotion probability is 83 p.p. (s.e. = 17) for women and 38 p.p. (s.e. = 13) for men, a difference that is marginally significant (p-value = 0.051). After 15 years the gap is still 57 p.p. (s.e.= 16) for women and 29 p.p. (s.e.= 12) for men. The absence of convergence likely reflects that unlucky candidates exited academia, moved abroad, or remained on non-tenured contracts — a pattern that appears particularly strong for women, consistent with failed qualification having especially severe consequences for their academic careers.

The picture is quite different for Full Professor candidates. Treatment effects decay steadily over time, becoming statistically insignificant after around six years, with point estimates close to zero by the end of the 15-year horizon. This pattern suggests that initially unsuccessful candidates catch up through reapplication. A likely contributing factor is that unsuccessful Full Professor candidates are typically already tenured Associate Professors, giving them a stable platform from which to reapply — in contrast to unsuccessful Associate Professor candidates, who may face greater uncertainty about their continued academic employment. At the Full Professor level, the effect appears less persistent for female than for male applicants, though the difference is not statistically significant, likely reflecting the small number of female Full Professor candidates in our sample (bottom panels in Figure 2).

6.1.2 Department-level evidence

We now examine whether individual qualification shocks propagate to the department level, using the 2SLS strategy described in Section 5.2. We begin by estimating the aggregate effect of qualification shocks on departmental outcomes, pooling across gender, before decomposing these effects by the gender of the qualifying candidate.

Mirroring the individual-level results, departments whose candidates draw more favorable committees experience significantly more qualifications. Appendix Table C5 shows that an additional connection between departments' candidates and the committee leads to 0.15 (s.e. = 0.01) additional qualifications in the department.

Turning to the second stage, each qualification leads to approximately one additional promotion at the corresponding rank in the short run, for both Associate and Full Professor positions (Appendix Figure F1). In the longer term, however, the pattern diverges across ranks. For Associate Professors, the cumulative effect grows steadily over time, while for Full Professors it modestly declines. Since department-level effects capture both the direct promotion of the qualifying candidate and any spillover effects on subsequent hiring, the growing Associate Professor effect is consistent with spillovers amplifying the direct effect over time. The fading Full Professor effect likely reflects the transitory nature of the individual-level shock, combined with the fact that unsuccessful Full Professor candidates typically remain in the department as tenured Associate Professors — meaning that a Full Professor qualification changes a candidate's rank but not their presence among tenured faculty.

We next examine whether the effects of Associate Professor qualifications differ by the gender of the qualifying candidate. We find that trickle-down effects are specific to female qualifications: only promoting a woman generates a lasting increase in female representation, without any offsetting effect on male promotions. More precisely, when a department's marginal female candidate exogenously qualifies to Associate Professor, the number of female Associate Professors in that department rises by about 0.6 (s.e. = 0.2) within one year (Figure 3, Panel A), mechanically reflecting the promotion of the qualifying candidate. The effect then increases gradually, reaching approximately 1.5 (s.e. = 0.8) additional female promotions after fifteen years. This suggests that, beyond the direct promotion of the qualifying candidate, exogenously promoting one woman may lead to approximately one additional female promotion in the long run. Appendix Figure F2 supports this interpretation using a specification that excludes promotions of any candidates from the department who had participated in the qualification exam: while this exclusion reduces precision, the point estimate remains close to one after 15 years, consistent with a substantial trickle-down effect.

While our main outcome is the flow of promotions, we can also examine the effect on the stock of tenured women in the department — a broader measure that captures not only promotions but also mobility and retirements. Appendix Figure F3 (Panel A) shows a positive effect on the stock of tenured women, though estimates are less precise than those for promotions, consistent with faculty stocks being affected by additional margins such as attrition, mobility, and retirement, which introduce noise beyond the promotion channel.

These effects are specific to the impact of women on women at the Associate Professor level. The qualification of a female candidate has no significant effect on male promotions at any horizon (Panel B of Figure 3), and male qualification shocks generate no significant long-run effects on either female or male promotions (Panels C and D).

Gender-specific results for Full Professor qualifications are reported in Appendix Figures F4 and F5; consistent with the aggregate results, we find no significant long-run effects, with estimates for female qualifications being particularly imprecise given the small number of female Full Professor candidates in our sample. Given the lack of persistent effects at the Full Professor rank, the remaining analysis focuses on Associate Professor qualifications.

6.2 Impact on PhD Graduates

We next examine whether the persistent change in faculty composition induced by female promotions translates into downstream effects on the number of female PhD graduates, their likelihood of remaining in academia, and their subsequent promotion prospects.

Figure 4 traces the cumulative effects of an exogenous female Associate Professor qualification on the number of PhD graduates. Departments that exogenously promote a woman train substantially more female PhD graduates over time (Panel A). An upward trend is already visible three to four years after the qualification, and effects reach statistical significance at around five years — broadly consistent with typical PhD program duration, though the early onset suggests that the effect is not driven solely by the recruitment of new students after the promotion shock. After ten years, departments have trained approximately six additional female PhD graduates in total. Note that these estimates capture the combined effect of the initial female qualification and any subsequent changes in departmental composition it generates, including additional female faculty hired as a result of the initial shock. The absence of any significant effect on male PhD graduates (Panel B) suggests that the increase reflects an expansion in cohort size rather than a reallocation away from male students.[19]

Female PhD graduates in affected departments are also substantially more likely to remain in academia. Figure 5 shows that approximately four additional female PhD graduates remain academically active following an exogenous female qualification. To illustrate the magnitude, consider the average department in our sample, which graduates roughly 13 female PhD students with a retention rate of 54% (Appendix Table A4). Following a female qualification shock, the same department would graduate approximately 19 female PhD students, of whom about 13 would remain academically active, raising the implied retention rate to 68%. Furthermore, Figure 6 shows that departments experiencing an exogenous female qualification produce one additional female PhD graduate who is promoted to Associate Professor within ten years of graduation. These effects are specific to female PhD graduates: female qualification shocks have no significant effect on the academic retention or promotion of male PhD graduates.

Turning to male qualification shocks, if anything these are associated with a modest reduction in female PhD graduates (see Panel C in Figures 5 and 6), though the effect is only marginally significant; male qualification shocks generate no significant effects on male PhD outcomes.

6.3 Heterogeneity across Fields

The aggregate estimates may conceal heterogeneity across fields. We examine two complementary dimensions: the baseline gender composition of the department and broad disciplinary groups (STEM+M versus SSH). These dimensions are related but not equivalent — female representation tends to be lower in STEM+M, yet many SSH departments also have very few senior women (Appendix Figure A2). STEM+M and SSH also differ in ways that may independently affect trickle-down dynamics, including the prevalence of team-based and lab research, gender stereotypes, and hiring practices. The representation split reveals whether trickle-down effects depend on the baseline presence of women, while the field split captures

19. Our data record PhD graduates rather than enrollees, so we cannot distinguish between whether this reflects the recruitment of additional female students or higher completion rates among existing ones.

broader differences in institutional structures and career paths.

We first estimate heterogeneous effects by baseline female representation in the department prior to the exam, measured as the share of women among Associate and Full Professors. We classify departments into three groups: (i) no senior women, (ii) women present but in the minority, and (iii) women in the majority. Panel A of Table 3 reveals pronounced heterogeneity by baseline female representation. In departments where women are present but in the minority (3,786 department-exam observations), a female qualification generates approximately two additional female Associate Professor promotions and eleven additional female PhD graduates within ten years — effects that are both large and precisely estimated. By contrast, effects are close to zero and statistically insignificant in departments with no senior women at baseline (2,013 observations), and positive but smaller and imprecisely estimated in departments where women are already in the majority (933 observations). While these coefficients are not statistically different, these results suggest that a minimal female presence is necessary for trickle-down effects to materialize: a lone woman may face too much isolation to generate spillovers, while departments that have already achieved parity see limited additional gains.

Panel B of Table 3 reports effects estimated ten years after the exam, separately for STEM+M and SSH departments. Effects are positive across all outcomes in both field groups and, while we cannot reject equality across fields, the point estimates do suggest some differences: trickle-down effects on female faculty promotions appear larger in STEM+M, while effects on female PhD graduates and their subsequent academic careers are somewhat stronger in SSH. Appendix Figure F6, which traces cumulative effects over a fifteen-year horizon separately for each field group, confirms that these patterns hold over the whole period.

Overall, the heterogeneity analysis points to baseline female representation as the key moderator: trickle-down effects are strongest where women are present but have not yet reached parity. Differences across disciplinary groups are less pronounced, with effects positive in both STEM+M and SSH, though the patterns within each group suggest that the mechanisms may operate somewhat differently across fields.

6.4 Gender and Research Orientation

Beyond promoting women directly, departments may also try to shift their gender composition by broadening hiring toward research areas that attract more gender-diverse student cohorts and future faculty. To examine this complementary channel, we estimate the effect of promoting candidates working in female-intensive research areas, independently of their gender. This exercise speaks to whether research-topic diversification — a policy that may face less institutional resistance than gender-targeted hiring — can generate downstream effects similar to those documented above. We then estimate a horse-race specification that includes both channels simultaneously to assess their relative importance.

Using the feminisation index described in Section 3, we classify candidates as working in *female-intensive* or *male-intensive* research areas depending on whether their index lies above or below the mean within their field (*área de conocimiento*) at the time of the exam. The average feminisation index is 30% for female candidates and 26% for male candidates, indicating meaningful but incomplete overlap: women are more likely to work in female-intensive areas but are not confined to them. In practice, 37% of candidates in female-intensive areas are female, compared to 28% in male-intensive areas. We apply the same department-level IV strategy to generate exogenous variation in the research orientation of qualified candidates.

Figure 7 shows that qualifying a candidate in a female-intensive research area generates persistent effects on departmental composition. Within six years, the number of promotions increases by 0.9 women (s.e. = 0.34). As shown in Appendix Figure F7, promotions of men increase by about 0.5 (s.e. = 0.44), reflecting the imperfect correlation between research topics and gender. In the long run, however, the effect becomes increasingly concentrated among women: departments gain on average 1.6 additional women (s.e. = 0.58) over fifteen years, compared to only 0.23 additional men (s.e. = 0.75). These are meaningful effects, comparable in magnitude to those generated by directly promoting a female candidate.

Turning to PhD outcomes, qualifying a candidate in a female-intensive research area also generates positive downstream effects, though somewhat smaller than those driven by directly promoting a woman. We estimate approximately four additional female PhD graduates over a ten-year horizon (Panel C, β=3.7, s.e.=2.0), of whom around two remain research-active (Panel E, β=1.8, s.e.=1.1) and about half go on to be promoted to Associate Professor within ten years of graduation (Panel G, β=0.49, s.e.=0.25). Appendix Figure F7 shows that the effects on male PhD outcomes are limited throughout. The contrast with the gender-based results suggests that research-topic alignment explains part of the effect on PhD outcomes, but that mechanisms specific to gender — such as mentoring, role models, and access to professional networks — are likely to play an important role as well.[20]

We next estimate a horse-race specification that includes both channels simultaneously: the number of qualified female candidates and the number of qualified candidates working in female-intensive research areas. Appendix Figure F9 reports the results for faculty promotions. The coefficient on female-intensive qualifications (Panel B) is positive and statistically significant at the 5% level: qualifying a candidate in a female-intensive research area, conditional on their gender, increases the number of female Associate Professor promotions by 1.2 over the subsequent 15 years (s.e. = 0.64). By contrast, the coefficient on female qualifications (Panel A) is close to one but imprecisely estimated.

Appendix Figure F10 reports the corresponding results for PhD outcomes. The effect of female qualifications on female PhD graduates remains large and stable even after controlling for female-intensive qualifications, whereas the corresponding coefficient for female-intensive qualifications is small and imprecise.

Taken together, these results suggest that broadening hiring toward female-intensive research areas can be an effective complementary policy for improving departmental gender composition, particularly at the faculty level. However, it does not fully replicate the effects of directly promoting women on PhD outcomes, where gender-specific mechanisms such as role models, mentoring, and departmental climate may also play an additional, quantitatively important role.

20. Appendix Figure F8 shows that qualifying a researcher working in non-feminised fields does not have any clear impact on female and male promotions and PhD graduates.

7 Interpretation and Mechanisms

The results documented above raise three related questions. First, why does failing to qualify for Associate Professor have such persistent consequences, and why are these consequences more severe for women than for men? Second, why does the exogenous promotion of a woman to Associate Professor generate spillovers on subsequent female faculty promotions? Third, why does it generate substantial downstream effects on female PhD graduates and their academic careers? We discuss each question in turn, before considering broader displacement and policy implications.

The cost of failing to qualify. Our individual-level evidence shows that failing to qualify in the *habilitación* exam has severe and persistent consequences, particularly for women. A marginally unsuccessful female candidate is around 83 p.p. less likely to be tenured ten years later than an otherwise similar candidate who qualified due to a favorable committee draw — a gap substantially larger than the corresponding 38 p.p. for men (Section 6.1.1). Several factors may explain why this bottleneck is especially consequential for women. First, unlike unsuccessful Full Professor candidates, who retain their Associate Professor tenure and can reapply, unsuccessful Associate Professor candidates face genuine uncertainty about their continued academic employment, and women may be less able to secure alternative positions or non-tenured contracts in the interim. Second, the *habilitación* period coincides with a stage of the career when many women face competing demands from family formation, making a failed qualification more likely to trigger a permanent exit from academia.

Spillovers on female faculty promotions. The finding that a single female promotion generates on average one additional female promotion over the subsequent fifteen years points to mechanisms that operate beyond the direct effect of the qualifying candidate. While the reduced-form estimates capture the overall downstream effect, they combine several dynamic components: the immediate promotion of the qualifying candidate, later promotions among initially unsuccessful candidates who eventually catch up, and potential trickle-down effects operating through the departmental environment. Appendix G combines the individual- and department-level estimates to provide a simple decomposition of these dynamics. The results suggest that much of the long-run departmental impact reflects gradual propagation through the departmental promotion pipeline, with spillovers emerging only several years after the initial promotion shock.

Several channels could account for this pattern — including direct mentoring and supervision of junior female researchers, role model and signaling effects that raise the perceived returns to an academic career among women, and shifts in departmental norms and hiring priorities. While we cannot directly distinguish between these channels, two pieces of evidence offer suggestive guidance.

First, the heterogeneity results indicate that effects are strongest where women are present but not yet in the majority, and close to zero in departments with no senior women at baseline. This pattern is more consistent with role model and signaling effects that require a critical mass than with a channel that operates from the moment the first woman enters a department.

Second, the research orientation results show that qualifying a candidate in a female-intensive research area generates long-run effects on female promotions of similar magnitude to those generated by qualifying a female candidate. This suggests that research agenda shapes hiring decisions at the faculty level, and that part of the gender effect on faculty composition operates through the research environment rather than through gender per se.

Downstream effects on PhD graduates. Our estimates combine potentially distinct margins — entry into PhD studies, completion, and subsequent persistence in academia. We do not observe applications, admissions, or enrollment, which prevents us from directly testing whether the promotion shock changes who enters a PhD program versus who completes it. However, the timing of effects provides useful evidence. We find increases in research activity and later promotions among female cohorts graduating as early as three to four years after the promotion shock. Given typical PhD durations, these cohorts were largely already enrolled when the promotion occurred, suggesting that part of the effect operates through changes affecting existing students — such as mentoring, improved departmental climate, or role model effects — rather than solely through the recruitment of new entrants.

We also assess whether the increase in female PhD graduates reflects changes in supervision patterns or broader departmental changes. Appendix Figure F11 shows that an additional female qualification leads to approximately two additional female PhD graduates supervised by women and about four supervised by men within ten years. These magnitudes suggest that the effect extends beyond direct supervision by the newly qualified woman, pointing instead to changes in departmental culture, climate, or research orientation that make the department more attractive or supportive for female doctoral students more broadly.

The contrast with the research orientation results is informative: qualifying a candidate in a female-intensive research area generates smaller and less precisely estimated effects on PhD outcomes than directly promoting a woman, suggesting that gender-specific mechanisms — such as role models and mentoring — play a more important role for doctoral students than research affinity alone.

Displacement. Our department-level estimates raise two related concerns about whether the observed increases in female participation reflect genuine expansion or reallocation. First, gains for women within a department could come at the expense of men. Second, increases in one department could be offset by reductions elsewhere, with our design unable to capture such cross-department displacement.

On the first concern, we find no evidence of within-department displacement: male promotion outcomes do not change significantly, and total promotion activity does not decline. Similarly, improvements in female PhD outcomes occur without significant changes in male PhD outcomes, ruling out a purely zero-sum adjustment within departments. On the second concern, cross-department reallocation is unlikely to be the main explanation for our findings, as both PhD recruitment and faculty mobility in Spain are predominantly local (Section 2).

8 Conclusions

Women's persistent underrepresentation in academia has motivated a wide range of policy interventions, often grounded in the belief that greater female representation at key career stages will generate broader benefits along the academic pipeline. Yet credible causal evidence on whether and how such trickle-down effects materialize has been scarce. This paper provides new evidence by exploiting quasi-random variation in promotions generated by Spain's centralized qualification system.

A first result concerns the individual cost of failing to obtain tenure. Among women, those who narrowly miss qualification are 83 percentage points less likely to be tenured ten years later than those who narrowly qualify, a gap that remains at 57 percentage points after fifteen years. This compares to gaps of 38 and 29 percentage points for men at the same horizons, suggesting that a failed qualification has especially severe and lasting consequences for women's academic careers. This finding reframes the policy debate: beyond asking how to generate trickle-down benefits from promoting women, it highlights the importance of preventing women from being permanently lost at the tenure stage in the first place.

When women do obtain tenure, the effects extend well beyond their own careers. Beyond the direct promotion of the qualifying candidate, departments experience on average one additional female promotion over the subsequent fifteen years, a substantial expansion of the female PhD pipeline, and improved retention and career progression among female doctoral graduates. These effects are concentrated in departments where women are underrepresented but not entirely absent, suggesting that a minimal female presence is a precondition for trickle-down effects to materialize, whether through mutual support, departmental influence, or other channels that require women not to be entirely isolated.

Research orientation plays an important and independent role alongside gender. Promoting researchers working in female-intensive areas generates downstream effects on hiring and PhD outcomes, indicating that part of what drives trickle-down effects operates through the topics, mentoring structures, and student demand associated with particular fields rather than through gender per se. At the same time, the effects of research orientation on PhD outcomes are substantially smaller than those driven by directly promoting a woman, pointing to gender-specific mechanisms — such as role models and mentoring — that research-topic diversification alone cannot replicate.

These findings have direct implications for the design of gender-targeted policies. Importantly, our compliers tend to be highly productive researchers, so our findings speak to policies that reduce promotion frictions for capable women at the margin, not to mechanically increasing representation or lowering standards. Our results provide direct causal support for affirmative action policies that raise the probability of promotion for strong female candidates who might otherwise be passed over: promoting a woman to Associate Professor generates large downstream benefits for future female faculty and for the PhD pipeline.

Policies that reduce the risk of such losses — whether by expanding promotion opportunities, providing bridge support to candidates facing temporary setbacks, or reducing the role of arbitrary factors in tenure decisions — may therefore be among the most effective tools for improving long-run female representation in academia.

More broadly, this paper shows how small changes at key career bottlenecks can reshape academic trajectories over long horizons. Promoting a single highly capable woman earlier in her career can set in motion a chain of effects that influences who enters academia, who remains, and who ultimately advances. Understanding these dynamic and cumulative effects is essential for designing policies that do not merely change who holds positions today, but who will shape the production of knowledge in the future.

References

Auriol, **Laudeline**, **Max Misu and Rebecca Ann Freeman**. 2013. "Careers of Doctorate Holders: Analysis of Labour Market and Mobility Indicators". *OECD Science, Technology and Industry Working Papers* 2013 (4).

Azmat, **Ghazala and Anne Boring**. 2020. "Gender Diversity in Firms". *Oxford Review of Economic Policy* 36 (4), pp. 760–782.

Bagues, **Manuel**, **Mauro Sylos-Labini and Natalia Zinovyeva**. 2017. “Does the Gender Composition of Scientific Committees Matter?” *American Economic Review* 107 (4), pp. 1207–38.

Bello, **Piera**, **Alessandra Casarico and Debora Nozza**. 2025. “Research Similarity and Women in Academia”. *The Economic Journal*, ueaf113.

Belot, **Michèle**, **Madina Kurmangaliyeva and Johanna Reuter**. 2023. *Gender Diversity and Diversity of Ideas*. Working Paper 16631. Institute of Labor Economics (IZA).

Bettinger, **Eric P and Bridget Terry Long**. 2005. “Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students”. *American Economic Review* 95 (2), pp. 152–157.

Borusyak, **Kirill and Peter Hull**. 2023. “Nonrandom Exposure to Exogenous Shocks”. *Econometrica* 91 (6), pp. 2155–2185.

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

Broder, **Ivy E**. 1993. “Review of NSF Economics Proposals: Gender and Institutional Patterns”. *The American Economic Review* 83 (4), pp. 964–970.

Burnham, **Judy F.** 2006. “Scopus Database: A Review”. *Biomedical Digital Libraries* 3 (1), p. 1.

Card, **David**, **Stefano DellaVigna**, **Patricia Funk and Nagore Iriberri**. 2020. “Are Referees and Editors in Economics Gender Neutral?” *The Quarterly Journal of Economics* 135 (1), pp. 269–327.

Carrell, **Scott E.**, **Marianne E. Page and James E. West**. 2010. “Sex and Science: How Professor Gender Perpetuates the Gender Gap”. *The Quarterly Journal of Economics* 125 (3), pp. 1101–1144.

Cruz Castro, **Laura**, **Luis Sanz Menendez and Jaime Aja Valle**. 2006. *Las Trayectorias Profesionales y Academicas de Los Profesores de Universidad y Los Investigadores Del CSIC*. Working Paper. Consejo Superior de Investigaciones Cientificas (Espana).

Cullen, **Zoë and Ricardo Perez-Truglia**. 2023. “The Old Boys’ Club: Schmoozing and the Gender Gap”. *American Economic Review* 113 (7), pp. 1703–40.

Deschamps, **Pierre**. 2024. “Gender Quotas in Hiring Committees: A Boon or a Bane for Women?” *Management Science* 70 (11), pp. 7486–7505.

Dolado, **Juan J.**, **Florentino Felgueroso and Miguel Almunia**. 2012. “Are Men and Women-Economists Evenly Distributed across Research Fields? Some New Empirical Evidence”. *SERIEs* 3 (3), pp. 367–393.

Epifanio, **Mariaelisa and Vera E. Troeger**. 2020. “Bargaining over Maternity Pay: Evidence from UK Universities”. *Journal of Public Policy* 40 (3), pp. 349–374.

European Commission. 2024. *She Figures 2024: Gender in Research and Innovation: Statistics and Indicators*. Tech. rep. Luxembourg: Publications Office of the European Union.

Folke, **Olle and Johanna Rickne**. 2022. "Sexual Harassment and Gender Inequality in the Labor Market*". *The Quarterly Journal of Economics* 137 (4), pp. 2163–2212.

Fuentes-Pujol, **Eulàlia and Llorenc Arguimbau-Vivo**. 2010. "Las Tesis Doctorales En Espana (1997-2008): Analisis, Estadisticas y Repositorios Cooperativos". *Revista espanola de documentacion cientifica* 33 (1), pp. 63–89.

Garrido Ardila, **Juan Antonio**. 2020. "Review of Dialnet. Database". *Renaissance and Reformation / Renaissance et Reforme* 43 (4), pp. 250–254.

Gaule, **Patrick and Mario Piacentini**. 2018. "An Advisor like Me? Advisor Gender and Post-Graduate Careers in Science". *Research Policy* 47 (4), pp. 805–813.

Ginther, **Donna K.**, **Janet M. Currie**, **Francine D. Blau and Rachel T. A. Croson**. 2020. "Can Mentoring Help Female Assistant Professors in Economics? An Evaluation by Randomized Trial". *AEA Papers and Proceedings* 110, pp. 205–09.

Giraneza Birekeraho, **Ariane and Paola Maniga**. 2018. *How Many Female Economics Professors in Top European Universities?* Blog Post. Bruegel.

Hafner, **Flavio and Christoph Hedtrich**. 2024. *Networks in the Market for Researchers*. Tech. rep. Working Paper.

Hilmer, **Christiana and Michael Hilmer**. 2007. "Women Helping Women, Men Helping Women? Same-Gender Mentoring, Initial Job Placements, and Early Career Publishing Success for Economics PhDs". *American Economic Review* 97 (2), pp. 422–426.

Hofstra, **Bas**, **Daniel A. McFarland**, **Sanne Smith and David Jurgens**. 2022. "Diversifying the Professoriate". *Socius* 8.

Hsieh, **Chang-Tai**, **Erik Hurst**, **Charles I. Jones and Peter J. Klenow**. 2019. "The Allocation of Talent and U.S. Economic Growth". *Econometrica* 87 (5), pp. 1439–1474.

Huang, **Junming**, **Alexander J. Gates**, **Roberta Sinatra and Albert-László Barabási**. 2020. "Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines". *Proceedings of the National Academy of Sciences* 117 (9), pp. 4609–4616.

Iaria, **Alessandro**, **Carlo Schwarz and Fabian Waldinger**. 2022. *Gender Gaps in Academia: Global Evidence Over the Twentieth Century*. SSRN Scholarly Paper. Rochester, NY.

Imbens, **Guido W. and Donald B. Rubin**. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models". *The Review of Economic Studies* 64 (4), pp. 555–574.

Kozlowski, **Diego**, **Vincent Larivière**, **Cassidy R. Sugimoto and Thema Monroe-White**. 2022. “Intersectional Inequalities in Science”. *Proceedings of the National Academy of Sciences* 119 (2), e2113067119.

Kunze, **Astrid and Amalia R. Miller**. 2017. “Women Helping Women? Evidence from Private Sector Data on Workplace Hierarchies”. *The Review of Economics and Statistics* 99 (5), pp. 769–775.

Lundberg, **Shelly and Jenna Stearns**. 2019. “Women in Economics: Stalled Progress”. *Journal of Economic Perspectives* 33 (1), pp. 3–22.

Mogstad, **Magne and Alexander Torgovitsky**. 2024. “Chapter 1 - Instrumental Variables with Unobserved Heterogeneity in Treatment Effects”. *Handbook of Labor Economics*. Ed. by **Christian Dustmann and Thomas Lemieux**. Vol. 5. Elsevier, pp. 1–114.

Moss-Racusin, **Corinne A.**, **John F. Dovidio**, **Victoria L. Brescoll**, **Mark J. Graham and Jo Handelsman**. 2012. “Science Faculty’s Subtle Gender Biases Favor Male Students”. *Proceedings of the National Academy of Sciences* 109 (41), pp. 16474–16479.

Neumark, **David and Rosella Gardecki**. 1998. “Women Helping Women? Role Model and Mentoring Effects on Female Ph.D. Students in Economics”. *The Journal of Human Resources* 33 (1), pp. 220–246.

Porter, **Catherine and Danila Serra**. 2020. “Gender Differences in the Choice of Major: The Importance of Female Role Models”. *American Economic Journal: Applied Economics* 12 (3), pp. 226–54.

Priem, **Jason**, **Heather Piwowar and Richard Orr**. 2022. *OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts.*

Rossello, **Giulia**, **Robin Cowan and Jacques Mairesse**. 2024. “Ph.D. Publication Productivity: The Role of Gender and Race in Supervision in South Africa”. *Journal of Productivity Analysis* 61 (3), pp. 215–227.

U.S. Department of Education. 2022. *National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Human Resources Component, Table 315.20.*

U.S. National Science Foundation. 2024. *Doctorate Recipients from U.S. Universities: 2023.* Tech. rep. NSF 25-300. Alexandria, VA: U.S.: National Center for Science and Engineering Statistics (NCSES).

Wang, **Kuansan**, **Zhihong Shen**, **Chiyuan Huang**, **Chieh-Han Wu**, **Darrin Eide**, **Yuxiao Dong**, **Junjie Qian**, **Anshul Kanakia**, **Alvin Chen and Richard Rogahn**. 2019. “A Review of Microsoft Academic Services for Science of Science Studies”. *Frontiers in Big Data* 2.

Wasserman, **Melanie**. 2023. “Gender Differences in Politician Persistence”. *The Review of Economics and Statistics* 105 (2), pp. 275–291.

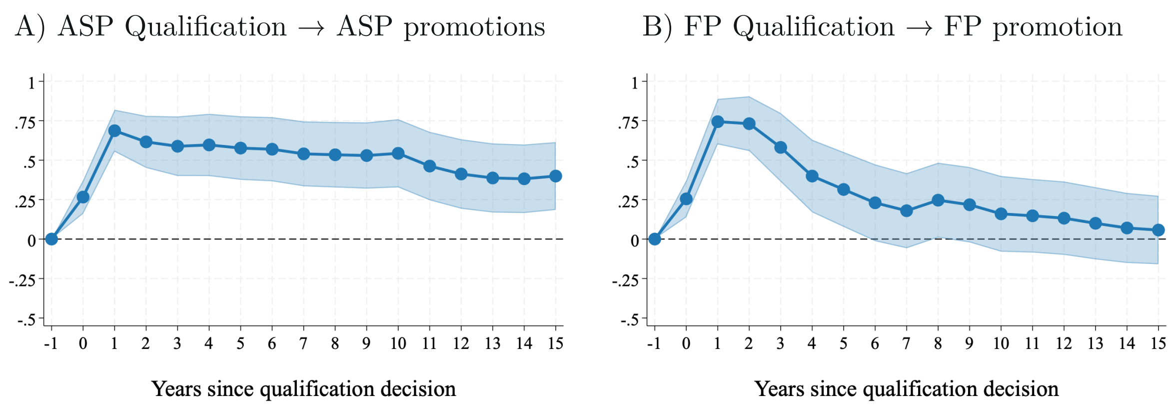
Williams, **Wendy M. and Stephen J. Ceci**. 2015. “National Hiring Experiments Reveal 2:1 Faculty Preference for Women on STEM Tenure Track”. *Proceedings of the National Academy of Sciences* 112 (17), pp. 5360–5365.

Wissenschaftskonferenz, **Gemeinsame**. 2016. *Pakt Für Forschung Und Innovation-Monitoring-Bericht. 2016*. Gemeinsame Wissenschaftskonferenz (GWK).

Zinovyeva, **Natalia and Manuel Bagues**. 2015. “The Role of Connections in Academic Promotions”. *American Economic Journal: Applied Economics* 7 (2), pp. 264–292.

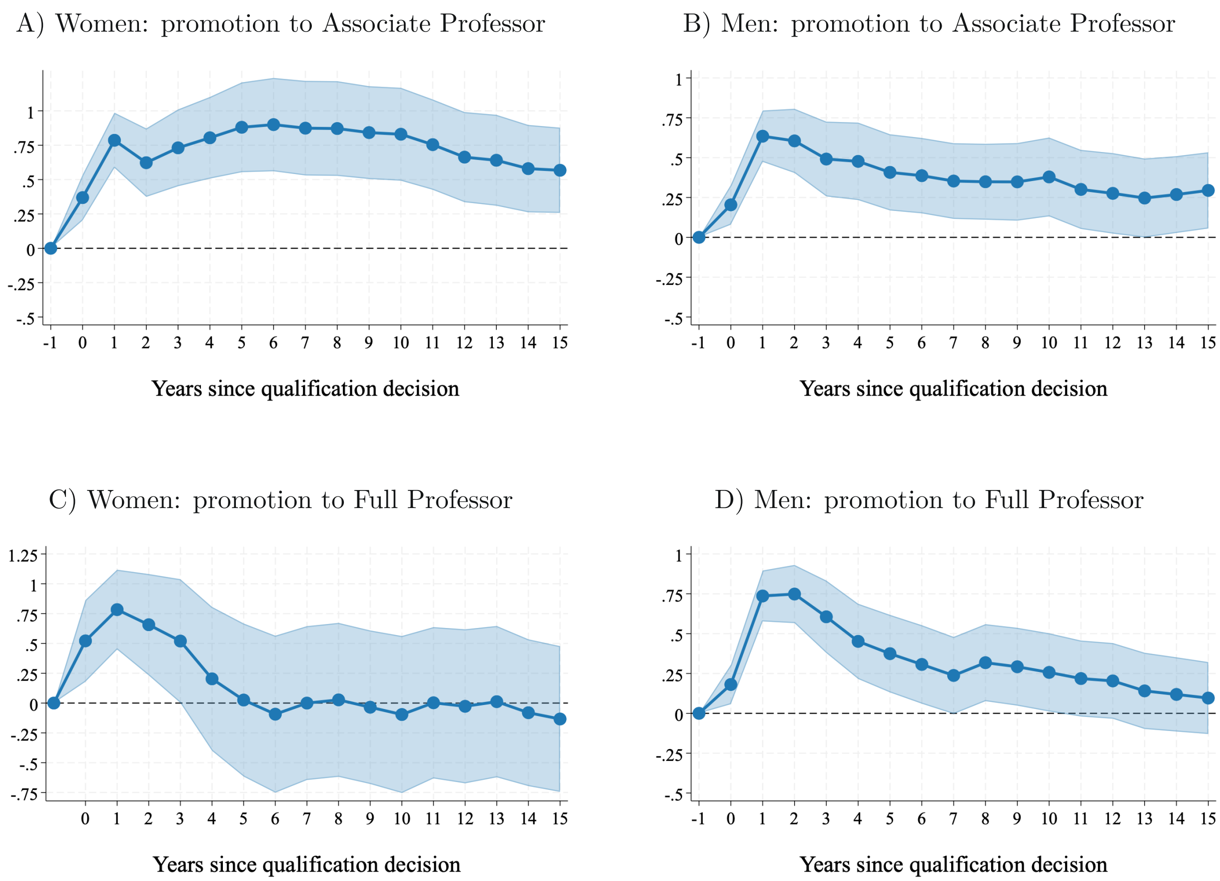
Figures

Figure 1: Individual-level effects of qualifications on *Associate* and *Full Professor* promotion



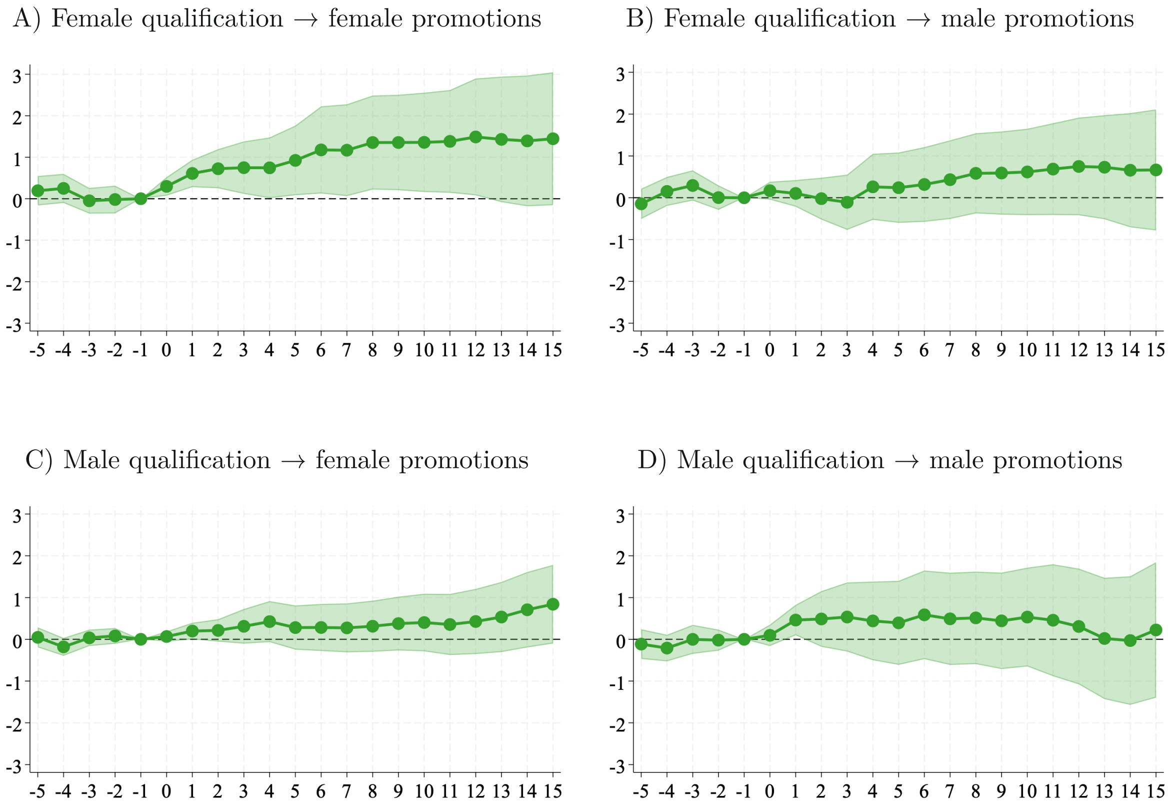
Notes: The figure reports 2SLS estimates of coefficient β_1 from Equation 2, capturing the impact of exogenous qualification in a given rank on the probability of promotion to that same rank in the years following the exam. Panel A reports effects on promotion to Associate Professor, while Panel B reports effects on promotion to Full Professor. The unit of analysis is the candidate–exam cell. Standard errors are clustered at the exam level. Shaded areas represent 95% confidence intervals.

Figure 2: Individual-level effects of qualification on promotion, by gender



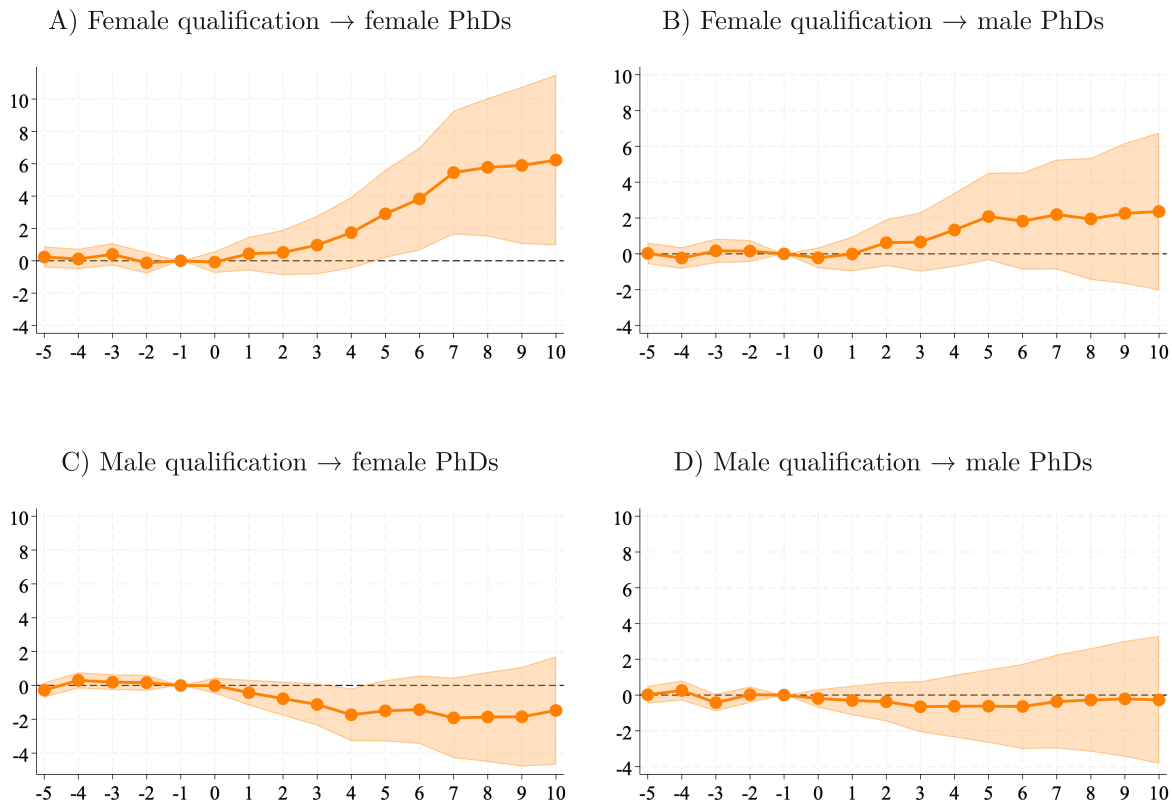
Notes: The figure reports 2SLS estimates of coefficient β_1 from Equation 2, capturing the impact of exogenous qualification in a given rank on the probability of promotion to that same rank in the years following the exam. Panels A and B report effects on promotion to Associate Professor for women and men, respectively. Panels C and D report effects on promotion to Full Professor for women and men, respectively. The unit of observation is the candidate–exam cell. Standard errors are clustered at the exam level. Shaded areas represent 95% confidence intervals.

Figure 3: Department-level effects of female and male qualifications to *Associate Professor* on promotions to *Associate Professor*



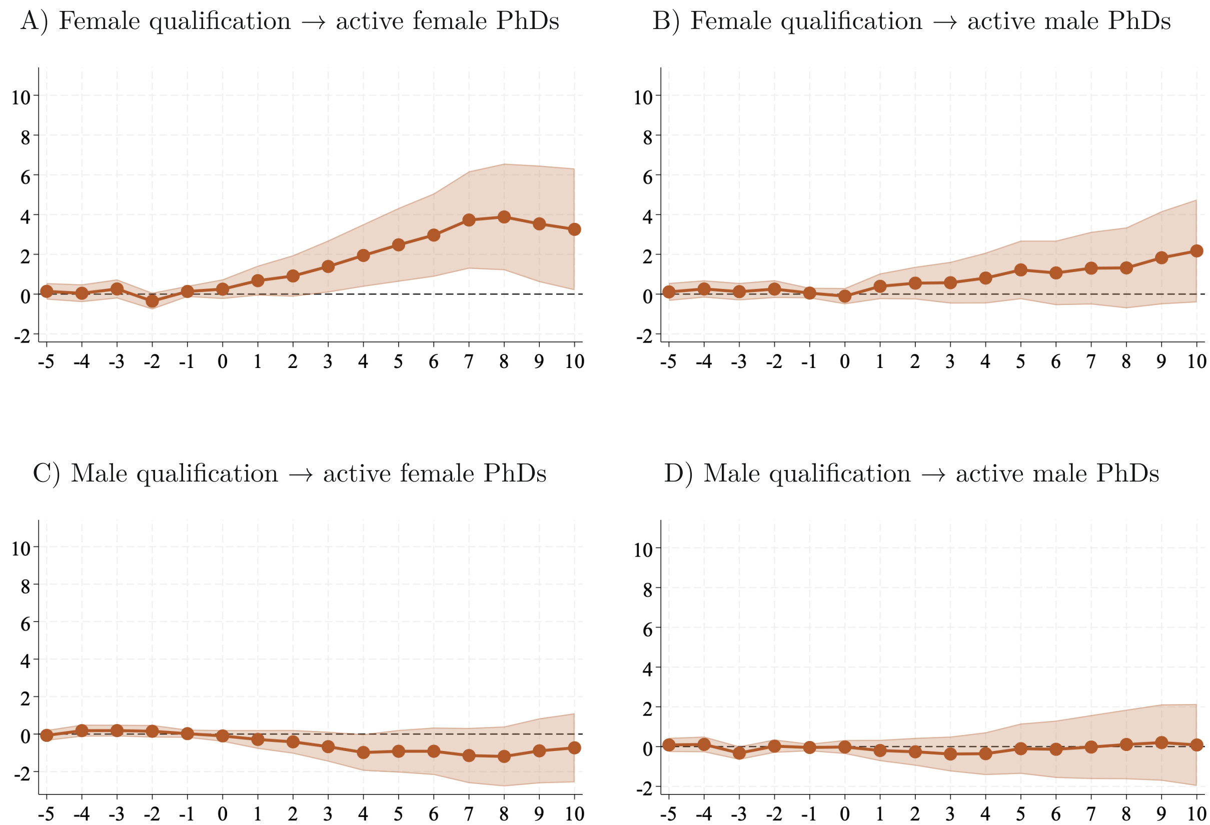
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6). Panels A and B show the effect of an exogenous female qualification to *Associate Professor* on the number of female and male promotions to *Associate Professor*, respectively. Panels C and D show the corresponding effects of an exogenous male qualification. Outcomes are measured relative to the year of qualification and are cumulative over time. The unit of analysis is the department–exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure 4: Department-level effects of female and male qualifications to *Associate Professor* on the number of PhD graduates



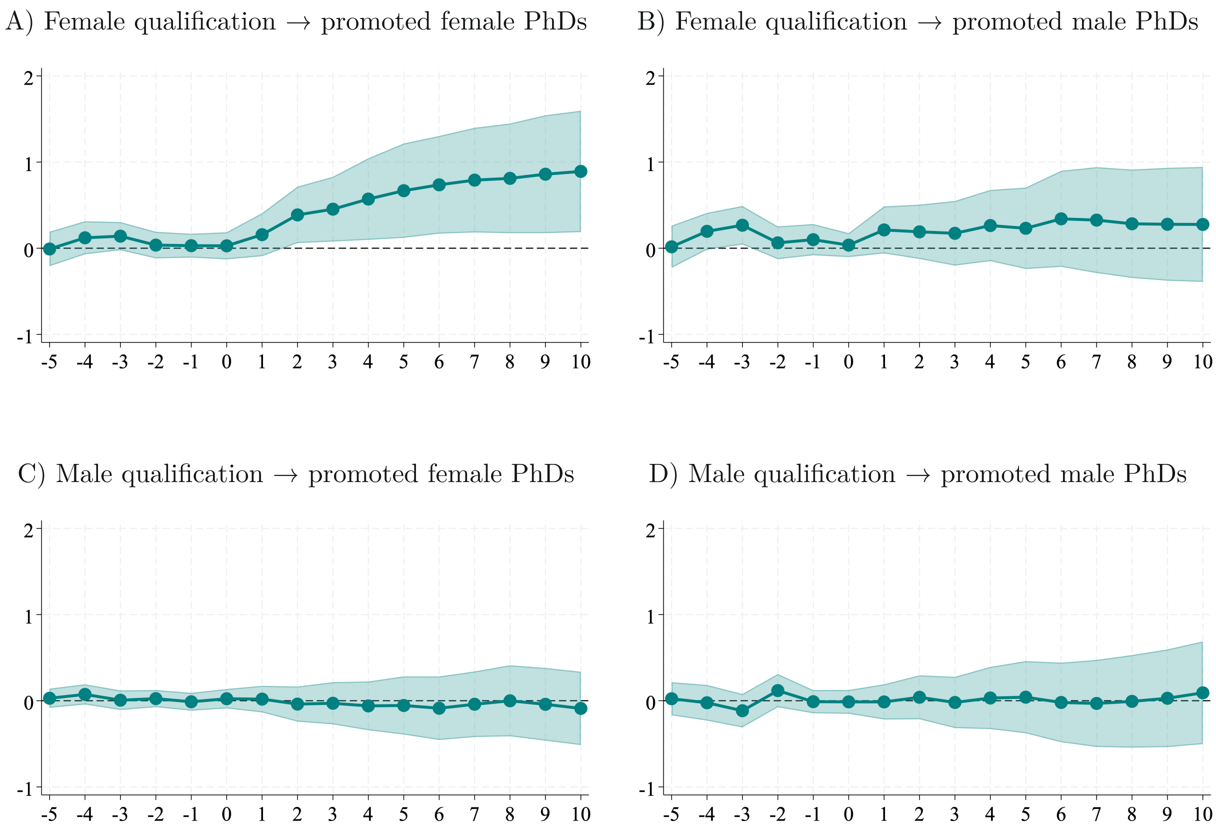
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6), capturing the impact of exogenously qualifying a woman or a man to Associate Professor on the number of female and male PhD graduates. Outcomes are measured relative to the year of qualification and are cumulative over time. Panels A and B show the effects of female qualifications, while Panels C and D show the effects of male qualifications. The unit of analysis is the department-exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure 5: Department-level effects of female and male qualifications to *Associate Professor* on the number of PhD graduates who remain research active 5-years post-graduation



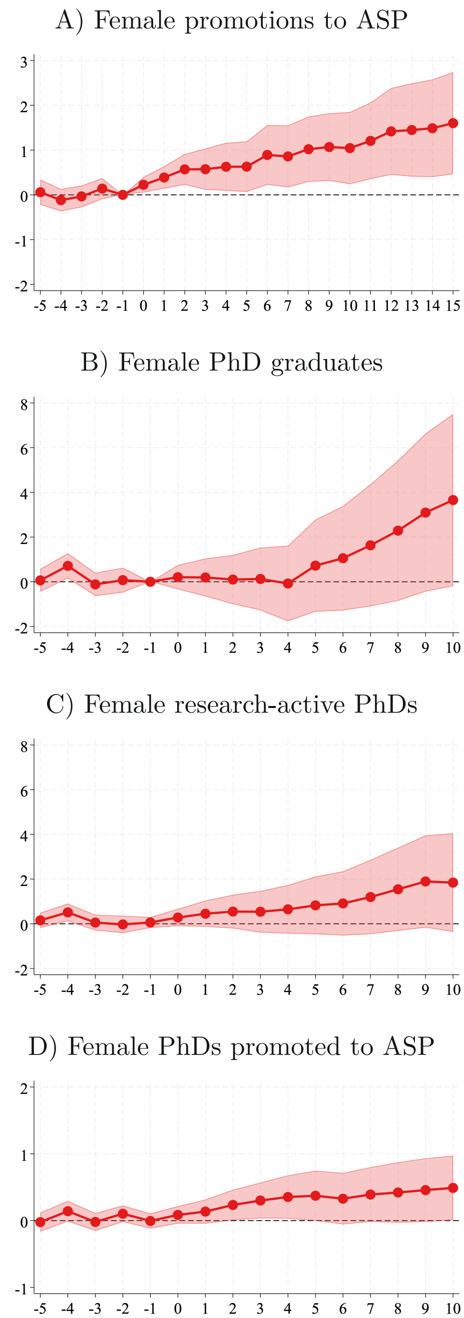
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6), capturing the impact of exogenously qualifying a woman or a man to *Associate Professor* on the number of female and male research-active PhD graduates – as measured by publishing 5 years post graduation. Outcomes are measured relative to the year of qualification and are cumulative over time. Panels A and B show the effects of female qualifications, while Panels C and D show the effects of male qualifications. The unit of analysis is the department–exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure 6: Department-level effects of female and male qualifications to *Associate Professor* on the number of PhD graduates who get promoted to Associate Professor within 10 years of graduation



Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6), capturing the impact of exogenously qualifying a woman or a man to *Associate Professor* on the number of female and male PhD graduates promoted to Associate Professor within 10 years since graduation. Outcomes are measured relative to the year of qualification and are cumulative over time. Panels A and B show the effects of female qualifications, while Panels C and D show the effects of male qualifications. The unit of analysis is the department–exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure 7: Department-level effects of *ASP* qualifications in female-intensive research areas



Notes: The figure reports 2SLS estimates of coefficient δ_1 from a specification analogous to Equation (6), where the endogenous variables are the number of qualified candidates working in female-intensive research areas and the number of those who do not. The coefficient captures the impact of exogenously qualifying such an applicant to *Associate Professor*. Each panel reports the dynamic effect of exogenously qualifying candidates working in more female-intensive research areas to *Associate Professor* on the outcome indicated in the panel title. Outcomes are measured relative to the year of qualification and are cumulative over time. The unit of observation is the department-exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Tables

Table 1: Senior female representation and gender composition of new entrants

Share of women among:	New PhDs (t+1)		New ASP (t+1)	
	(1)	(2)	(3)	(4)
Share Female FP	0.028***	0.028***	0.048***	0.047***
	(0.011)	(0.011)	(0.015)	(0.015)
Share Female ASP	0.038***	0.037***	0.082***	0.080***
	(0.013)	(0.013)	(0.018)	(0.018)
FP feminised research (std)		0.007		0.009
		(0.005)		(0.007)
ASP feminised research (std)		0.010*		0.021***
		(0.006)		(0.008)
Number of Observations	21536	21536	14070	14070
Number of Departments	3258	3258	3230	3230
Baseline	0.48	0.48	0.37	0.37
Adjusted R^2	0.128	0.128	0.109	0.110
Year FE	✓	✓	✓	✓
Field FE	✓	✓	✓	✓

Notes: OLS estimates. The unit of observation is a department–year. In columns (1)–(2), the outcome is incoming PhD students, inferred from PhD graduations observed through 2017; the sample is restricted to 2000–2011 to ensure that all entry cohorts are fully observed in the graduation data. In columns (3)–(4), the outcome is incoming Associate Professors, using data from 2000–2022. Standard errors are clustered at the department level. *FP* and *ASP* denote Full and Associate Professors. *FP/ASP feminised* research measures the gender composition of the research areas in which the average FP or ASP specializes, normalized to mean zero and unit standard deviation.
*** $p < .01$, ** $p < .05$, * $p < .1$

Table 2: PhD students' gender and the characteristics of their advisor

	Female student			
	(1)	(2)	(3)	(4)
Female Advisor	0.099***		0.092***	0.098***
	(0.004)		(0.004)	(0.005)
Advisor's research feminisation (std)		0.039***	0.032***	0.038***
		(0.003)	(0.003)	(0.003)
Observations	126156	126156	126156	126156
Number of departments	4006	4006	4006	4006
Baseline	0.43	0.43	0.43	0.43
R^2	0.122	0.119	0.124	0.320
Year FE	✓	✓	✓	
Department FE	✓	✓	✓	
Department × Year FE				✓

Notes: OLS estimates. The unit of analysis is a PhD student. The sample period is 1990–2017. Departments are university-field interactions. Standard errors clustered at department level. *Advisor's Research Feminization* is based on pre-graduation publications, and it is standardised to have zero mean and unit standard deviation.
*** p < .01, ** p < .05, * p < .1

Table 3: Impact of female qualification by baseline female representation and field group

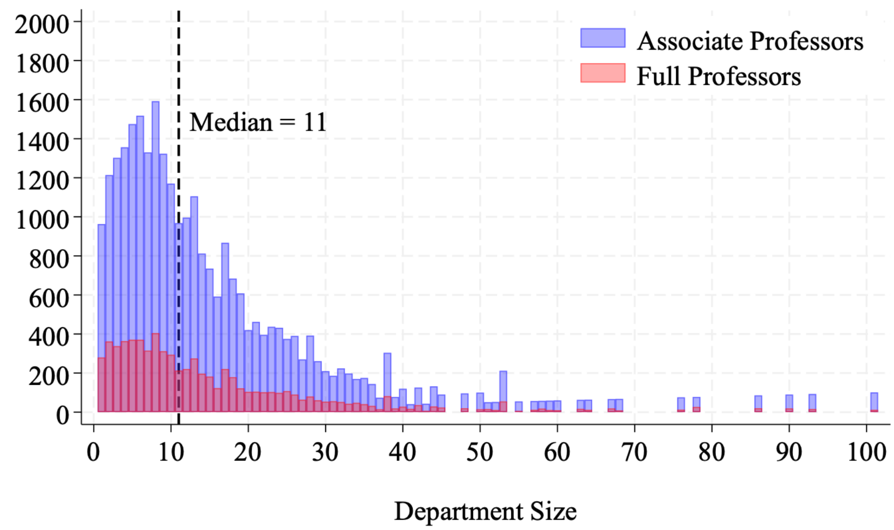
	Female promotions to Associate Professor (10 years)	Female PhD graduates (10 years)	Active Female PhD graduates (10 years)	Female PhD graduates promoted ASP (10 years)
Panel A. Heterogeneity by baseline number of women				
No senior women	0.116	-5.921	-2.623	0.566
[N= 2,013]	(0.625)	(4.435)	(2.276)	(0.457)
Women in minority (≤ men)	2.037**	10.999**	4.745**	0.993**
[N= 3,786]	(0.849)	(4.330)	(2.402)	(0.473)
Women in majority (> men)	1.037	6.776	5.037	0.931
[N= 933]	(1.246)	(4.477)	(3.068)	(0.801)
R^2	0.17	0.54	0.50	-0.02
Panel B. Heterogeneity by field				
SSH	0.884	7.023**	3.316*	0.761*
[N= 3,214]	(0.706)	(3.065)	(1.696)	(0.419)
STEMM	2.130**	4.807	3.089	1.122*
[N= 3,518]	(1.025)	(4.731)	(2.866)	(0.596)
R^2	0.17	0.57	0.53	-0.02

Notes: The table reports 2SLS estimates at the department–exam level for applications to Associate Professor positions. The outcome variables are the cumulative number of (i) female promotions to Associate Professor, (ii) female PhD graduates, (iii) research-active female PhD graduates, and (iv) female PhD graduates promoted to Associate Professor within 10 years from graduation. All outcomes are measured within ten years following the exam. Female qualification is instrumented using random variation in committee composition, as described in Section 5.2. All specifications include controls listed in Section 5.2. Panel A splits the sample by baseline female representation, measured one year prior to the exam using the combined pool of Full and Associate Professors. Departments are classified into three mutually exclusive categories: (i) no senior women; (ii) women in the minority, i.e., the number of senior women is weakly below the number of senior men; and (iii) women in the majority, i.e., the number of senior women is strictly greater than the number of senior men. Panel B splits the sample by field group: *SSH* includes Social Sciences and Humanities while *STEMM* includes STEM fields and Medicine.
*** p < .01, ** p < .05, * p < .1

Appendix A — Descriptive statistics

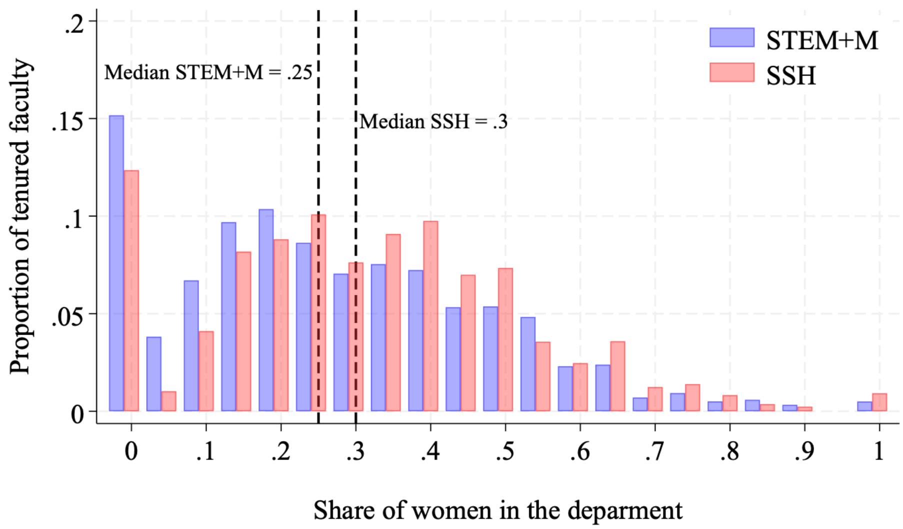
This appendix provides supplementary descriptive evidence on the structure of Spanish academia and the study sample. It documents the distribution of department size and gender composition, trends in female representation by rank, the distribution of the feminization index across researcher types, and summary statistics for PhD graduates, qualification exam candidates, and departments.

Figure A1: Distribution of Tenured Professors over Department Size



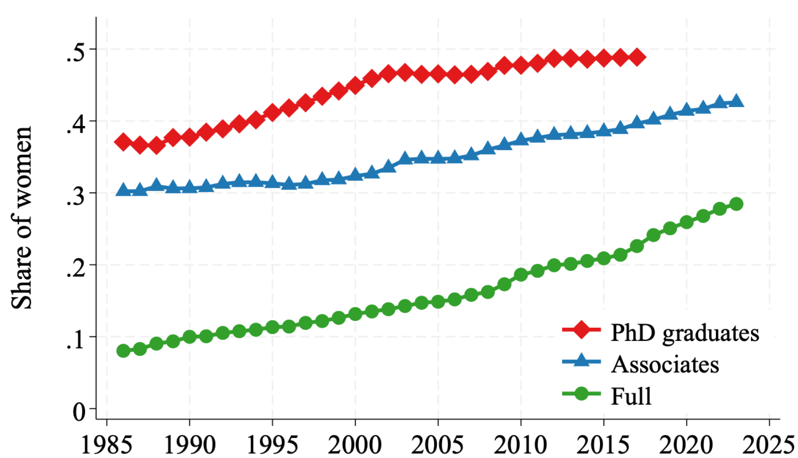
Notes: The figure displays the distribution of tenured faculty by department size in 2000. The histogram separately reports Associate Professors and Full Professors, with the vertical dashed line indicating the median department size (11 faculty members). The underlying data are drawn from the Spanish State Bulletin (BOE), based on our own calculations.

Figure A2: Distribution of tenured faculty by the share of women in the department



Notes: The figure displays the distribution of tenured faculty by the departmental share of women in 2000. The histogram separately reports STEM+Medicine and Social Sciences and Humanities (SSH) departments. The vertical dashed lines indicate the median departmental share of women in each broad field (0.25 in STEM+M and 0.30 in SSH), where the median is defined as the department employing the median tenured professor. The underlying data are drawn from the Spanish State Bulletin (BOE), based on our own calculations.

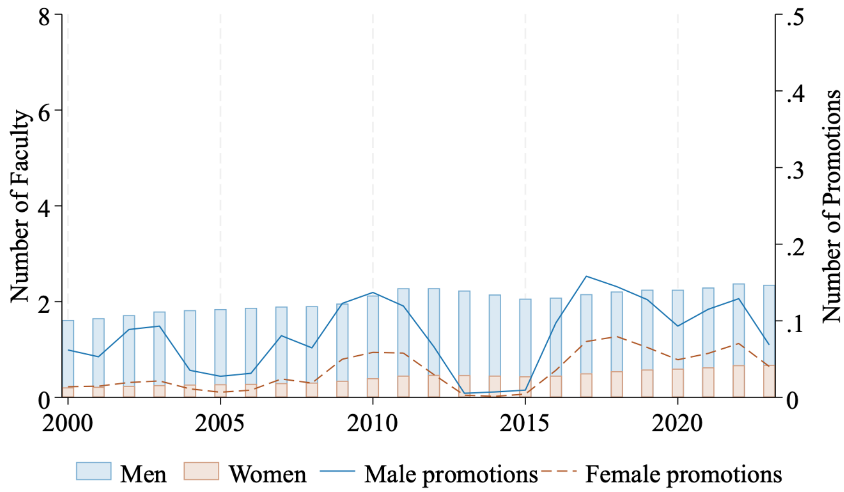
Figure A3: Share of women by academic role over time



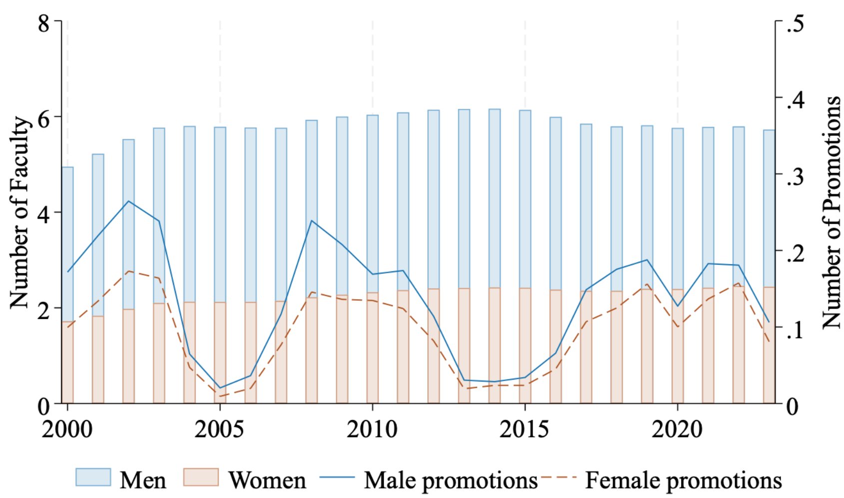
Notes: The figure displays trends in the share of women by academic role. The series report the fraction of women among PhD graduates (1985–2017) and among Associate and Full Professors (1985–2023), by year. Data are from PhD theses, appointment records of faculty eligible to serve on qualification exams, and academic promotions to Associate and Full Professor; our own calculations.

Figure A4: Annual promotion flows and stocks of Associate and Full Professors

A) Full Professors

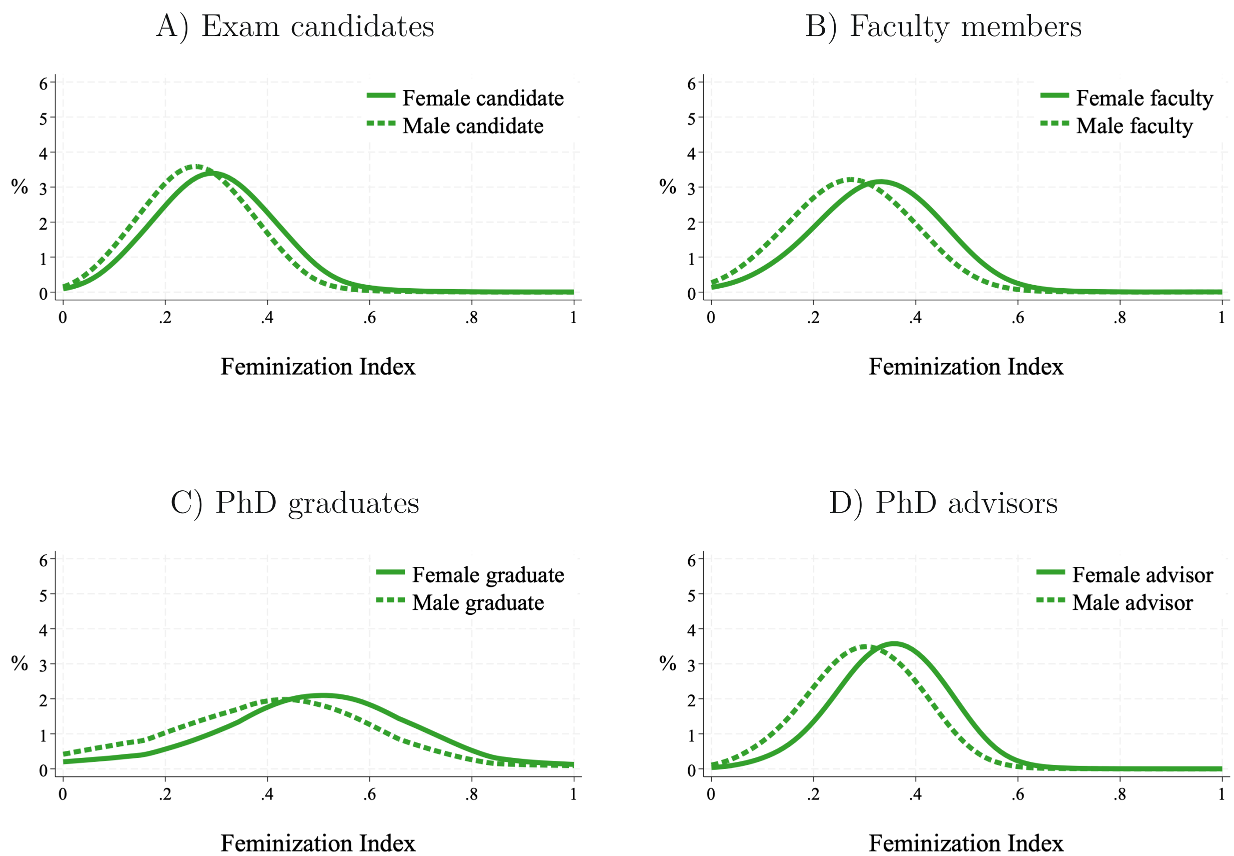


B) Associate Professors



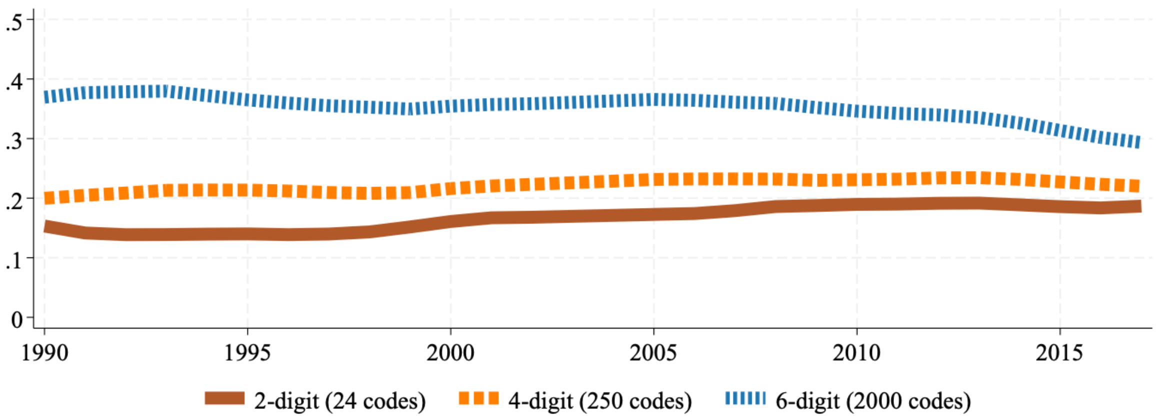
Notes: The figure shows annual promotions (lines) and the corresponding stocks (bars) of Full Professors (Panel A) and Associate Professors (Panel B) in Spanish public universities over time. Our own calculations using data from the Spanish State Bulletin (BOE).

Figure A5: Distribution of the feminization index by researcher type and gender



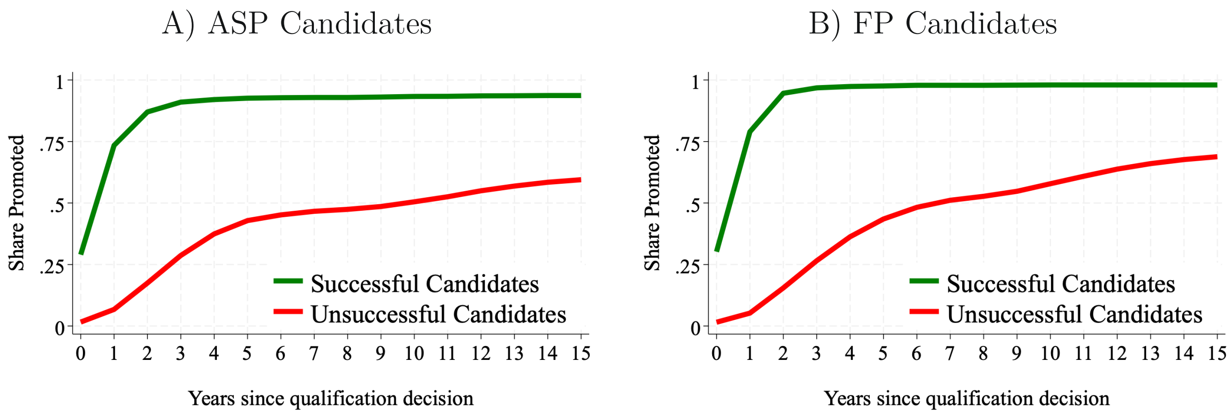
Notes: The figure shows the distribution of the author-level feminization index by gender, separately for exam candidates (Panel A), faculty members (Panel B), PhD graduates (Panel C), and PhD advisors (Panel D). The index measures how female-dominated the research topics of the average paper of an author are, calculated using the gender composition of research topics at the time of publication. See Section 3 for the construction of the index and Appendix H.5 for the mapping of the OpenAlex taxonomy to Teseo, Scopus, and Dialnet. Kernel density estimates use a bandwidth of 0.07 and 100 evaluation points.

Figure A6: Gender segregation in research topics among PhD graduates over time



Notes: The figure plots the Duncan segregation index for research topics between female and male PhD students over time, using the UNESCO taxonomy at three levels of granularity. The index measures the share of one group that would need to be reassigned across research fields to match the distribution of the other group. The sample includes all PhD students graduating from Spanish public universities between 1990 and 2017. The years 2003 and 2004 are excluded due to missing information on research topics in TESEO dataset.

Figure A7: Promotion rates by qualification outcome and years since exam



Notes: The figure plots the cumulative share of applicants who have been promoted, separately by qualification outcome, for candidates applying to Associate Professor positions (Panel A) and to Full Professor positions (Panel B). Outcomes are measured relative to the year of the qualification exam.

Table A1: Characteristics of PhD graduates in Spain

	Overall (1)	Men (2)	Women (3)	Women–Men (4)	SE (5)
Panel A: Overall [N= 203,542]					
Female Student	0.47				
Share Female Supervisors	0.21	0.16	0.27	0.11***	0.00
Years to Graduate	5.16	5.14	5.19	0.06***	0.01
Staying in Academia	0.53	0.54	0.52	-0.02***	0.00
Promoted ASP	0.10	0.12	0.08	-0.04***	0.00
Panel B: STEM+M Fields [N= 124,911]					
Female Student	0.47				
Share Female Supervisors	0.20	0.15	0.25	0.10***	0.00
Years to Graduate	4.89	4.87	4.91	0.04***	0.01
Staying in Academia	0.54	0.56	0.52	-0.04***	0.00
Promoted ASP	0.09	0.12	0.06	-0.06***	0.00
Accumulated research during PhD					
Papers	1.01	0.04	-0.04	-0.08***	0.01
Papers in Web of Science	0.68	0.02	-0.03	-0.05***	0.01
Total Article Influence Score	0.81	0.02	-0.02	-0.04***	0.01
Average Article Influence Score	1.14	0.01	-0.01	-0.01	0.01
Accumulated research 1-5 years after graduation (conditional on staying in academia)					
Papers	1.76	0.40	0.31	-0.10***	0.02
Papers in Web of Science	1.19	0.36	0.30	-0.07***	0.02
Total Article Influence Score	1.42	0.33	0.28	-0.05**	0.02
Average Article Influence Score	1.15	0.01	0.03	0.03	0.02
Panel C: SSH Fields [N= 78,631]					
Female Student	0.47				
Share Female Supervisors	0.24	0.18	0.29	0.11***	0.00
Years to Graduate	5.59	5.55	5.65	0.10***	0.01
Staying in Academia	0.51	0.51	0.52	0.01**	0.00
Promoted ASP	0.12	0.13	0.12	-0.02***	0.00
Accumulated research during PhD					
Papers	0.44	0.02	-0.03	-0.05***	0.01
Papers in Web of Science	0.08	0.01	-0.01	-0.02**	0.01
Total Article Influence Score	0.07	0.01	-0.01	-0.01	0.01
Average Article Influence Score	0.78	0.01	-0.01	-0.03	0.04
Accumulated research 1-5 years after graduation (conditional on staying in academia)					
Papers	0.86	0.34	0.29	-0.05	0.03
Papers in Web of Science	0.20	0.21	0.25	0.05	0.05
Total Article Influence Score	0.20	0.19	0.24	0.05	0.05
Average Article Influence Score	0.80	0.01	0.05	0.04	0.09

Notes: The unit of observation is a PhD graduate from a Spanish public university between 1990 and 2017. Panel A includes all graduates; Panels B and C restrict the sample to STEM+Medicine and Social Sciences and Humanities (SSH), respectively. Column (1) reports unconditional means. Columns (2) and (3) report means separately for women and men; for productivity outcomes, these are standardized within area-by-graduation-year cells to have mean zero and unit standard deviation, while all other variables are reported in levels. Column (4) reports the difference in means (women minus men) and Column (5) the corresponding standard error.
*** $p < .01$, ** $p < .05$, * $p < .1$

Table A2: Characteristics of candidates in qualification exams

	Overall (1)	Men (2)	Women (3)	Women–Men (4)	SE (5)
Panel A: Applications for Full Professor [N= 13,444]					
Female Candidate	0.27				
Qualified	0.11	0.11	0.09	-0.02***	0.01
Papers	33	0.05	-0.13	-0.18***	0.02
Papers in Web of Science	17	0.02	-0.05	-0.07***	0.02
Books	3	0.05	-0.13	-0.18***	0.02
Chapters	7	0.04	-0.11	-0.15***	0.02
Total AIS	17	0.02	-0.05	-0.07***	0.02
Average AIS	0.83	0.01	-0.03	-0.04*	0.02
Any strong tie	0.351	0.352	0.348	-0.004	0.009
Any colleague	0.322	0.322	0.323	0.001	0.009
Any PhD advisor	0.032	0.033	0.030	-0.003	0.003
Any coauthor	0.071	0.073	0.065	-0.008	0.005
Strong ties (intensity)	4.013	4.092	3.800	-0.292**	0.135
Panel B: Applications for Associate Professor [N= 17,799]					
Female Candidate	0.40				
Qualified	0.12	0.12	0.11	-0.01**	0.00
Papers	17	0.10	-0.15	-0.25***	0.01
Papers in Web of Science	8	0.07	-0.11	-0.17***	0.01
Books	1	0.06	-0.09	-0.15***	0.01
Chapters	3	0.02	-0.03	-0.05***	0.01
Total AIS	10	0.06	-0.10	-0.16***	0.01
Average AIS	0.87	0.03	-0.05	-0.08***	0.02
Any strong tie	0.292	0.295	0.286	-0.009	0.007
Any colleague	0.275	0.278	0.270	-0.009	0.007
Any PhD advisor	0.022	0.022	0.023	0.002	0.002
Any coauthor	0.039	0.039	0.039	0.000	0.003
Strong ties (intensity)	3.068	3.140	2.958	-0.181**	0.089

Notes: The unit of observation is an application to the *habilitación* qualification system. Panel A reports applicants to Full Professor positions; Panel B reports applicants to Associate Professor positions. Column (1) reports overall means. Columns (2) and (3) report means by gender; for productivity measures, these are standardized at the exam level, while all other variables are reported in levels. Column (4) reports the difference in means (women minus men) and Column (5) the corresponding standard error.
*** $p < .01$, ** $p < .05$, * $p < .1$

Table A3: Department participation in qualification exams

	Count	Percent
Departments in Spain	5422	
Departments with any Candidate in Habilitación	4000	100%
Panel A: Full professor exams		
Departments with:		
a candidate	2642	66%
a successful candidate	955	24%
a female candidate	1159	29%
a successful female candidate	286	7%
a male candidate	2304	58%
a successful male candidate	791	20%
Panel B: Associate professor exams		
Departments with:		
a candidate	3330	83%
a successful candidate	1224	31%
a female candidate	2080	52%
a successful female candidate	564	14%
a male candidate	2550	64%
a successful male candidate	866	22%

Notes: The table shows the distribution of Spanish public university departments by participation and success in *habilitación* exams. The sample includes all departments with at least one Full Professor and one Associate Professor. Percentages are computed relative to the 4,000 departments with at least one candidate. Panel A reports counts for Full Professor exams and Panel B for Associate Professor exams. "Departments with a candidate" denotes at least one applicant to the relevant exam; "departments with a successful candidate" denotes at least one qualified applicant. Gender-specific rows indicate the presence of at least one female or male candidate, and at least one successful female or male candidate, respectively.

Table A4: Characteristics of departments participating in qualification exams

	Departments with Applicants to Full Professor					Departments with Applicants to Associate Professor				
	Mean	SD	p10	p50	p90	Mean	SD	p10	p50	p90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Department-exam level characteristics										
Faculty participating in the exam										
Female	0.56	0.87	0	0	2	0.93	1.25	0	1	2
Male	1.50	1.44	0	1	3	1.42	1.77	0	1	3
Faculty qualified in the exam										
Female	0.05	0.23	0	0	0	0.11	0.36	0	0	0
Male	0.17	0.42	0	0	1	0.18	0.46	0	0	1
Tenured faculty 1 year prior to the exam										
Female	3.92	4.84	0	2	10	3.29	4.60	0	2	9
Male	8.61	9.42	1	6	18	7.68	9.69	0	5	18
Panel B. Cumulative outcomes 10 years post-exam										
Promotions to Full Professor										
Female	0.71	1.24	0	0	2	0.48	1.00	0	0	2
Male	1.74	2.08	0	1	4	1.33	1.94	0	1	4
Promotions to Associate Professor										
Female	1.36	2.10	0	1	4	1.49	2.17	0	1	4
Male	1.98	3.16	0	1	5	2.36	3.51	0	1	6
PhD graduates										
Female	14.12	16.24	1	8	37	12.82	15.34	1	7	34
Male	14.92	16.05	1	9	38	14.26	15.84	1	9	37
PhD graduates staying active in academia										
Female	7.56	8.84	0	4	20	6.92	8.32	0	4	18
Male	7.96	8.81	0	5	20	7.63	8.71	0	5	20
PhD graduates promoted to Associate Professor										
Female	0.46	0.94	0	0	2	0.53	1.05	0	0	2
Male	0.79	1.63	0	0	2	0.91	1.79	0	0	3
Number of department-exam			6485					6733		
Number of departments			2642					3330		

Notes: The table reports department–exam level descriptive statistics for departments in Spanish public universities with at least one faculty member applying to a national qualification exam. Columns (1)–(5) cover departments participating in Full Professor exams; columns (6)–(10) cover Associate Professor exams. Panel A reports counts of candidates and tenured faculty at the time of the exam. Panel B reports cumulative promotions and PhD graduates over the ten years following the exam.

Appendix B — Randomization Procedure and Expected Connectedness

This appendix describes the random assignment procedure used to select evaluation committees and derives the expected number of connections between candidates and committee members under this procedure.

Randomization Procedure

Committee members were selected by random draw from the list of eligible evaluators in the relevant field. Eligible evaluators were professors and researchers working in a public institution in Spain at the time of the exam, who met a minimum research quality threshold based on the number of *sexenios* — periods of research activity recognized by the Spanish education authorities.[21] The selection was carried out by Ministry officials using a drum containing one ball per eligible evaluator.

Each committee consisted of seven members. In Associate Professor exams, three evaluators were drawn from the list of eligible Full Professors and four from the list of eligible Associate Professors. In Full Professor exams, all seven members were drawn from the list of eligible Full Professors. The committee member with the longest tenure was appointed president, and the exam was held at the university where the president was based. A reserve committee of seven evaluators was also randomly drawn to replace any resignations. Evaluators could resign only under a restricted set of circumstances, and resignations were rare: approximately 2% of initially assigned evaluators were replaced.

The random assignment was subject to two minor constraints: no more than one researcher from the Spanish Research Council (CSIC) and no more than one emeritus professor could

21. Approximately 80% of Full Professors and 70% of Associate Professors met this requirement (*Comisión Nacional Evaluadora de la Actividad Investigadora*, Memoria de los resultados de las evaluaciones realizadas de 1989 a 2005, 2005, available via Web Archive at http://www.mecd.gob.es/dctm/ministerio/horizontales/ministerio/organismos/cneai/2005-memoria-1989-2005-universidad.pdf?documentId=0901e72b8008d9f6).

serve on a given committee.[22] When a second individual in either category was drawn, the draw was repeated. These constraints are taken into account when computing each candidate's expected connectedness, as described below.

Computing Expected Connectedness

Given the randomization procedure, we compute the expected number of connections $\mathbb{E}[C_{i,e}]$ for each candidate analytically, accounting for the CSIC and emeritus constraints.

Full Professor exams. In Full Professor exams, all seven committee members are drawn from the pool of eligible Full Professors. Let R, E, and P denote the number of researchers (CSIC), emeritus professors, and other eligible professors, respectively. The probability that at least one researcher and at least one emeritus professor is drawn is

$$p_R = 1 - \frac{\binom{R}{0}\binom{P+E}{7}}{\binom{P+E+R}{7}}, \qquad p_E = 1 - \frac{\binom{E}{0}\binom{P+R}{7}}{\binom{P+E+R}{7}}.$$

The expected number of connections is then:

$$\begin{aligned}\mathbb{E}[C_{i,e}] &= p_R p_E(c_R + c_E + 5c_P) + p_E(1 - p_R)(c_E + 6c_P) \\ &\quad + p_R(1 - p_E)(c_R + 6c_P) + (1 - p_R)(1 - p_E) \cdot 7c_P,\end{aligned}$$

where c_j denotes candidate i's number of connections with members of group $j \in \{R, E, P\}$.

Associate Professor exams. In Associate Professor exams, three evaluators are drawn from the pool of eligible Full Professors and four from the pool of eligible Associate Professors. We compute the probabilities p_R^{FP}, p_E^{FP}, p_R^{AP}, and p_E^{AP} analogously to the Full Professor case,

22. These constraints affect 387 of the 967 exams in our sample.

separately for each pool. The expected number of connections is:

$$
\begin{aligned}
\mathbb{E}[C_{i,e}] = & \Big[p_R^{FP} p_E^{FP} (c_R^{FP} + c_E^{FP} + c_P^{FP}) + p_E^{FP}(1 - p_R^{FP})(c_E^{FP} + 2c_P^{FP}) \\
& + p_R^{FP}(1 - p_E^{FP})(c_R^{FP} + 2c_P^{FP}) + (1 - p_R^{FP})(1 - p_E^{FP}) \cdot 3c_P^{FP}\Big] \\
& + \Big[p_R^{AP} p_E^{AP} (c_R^{AP} + c_E^{AP} + 2c_P^{AP}) + p_E^{AP}(1 - p_R^{AP})(c_E^{AP} + 3c_P^{AP}) \\
& + p_R^{AP}(1 - p_E^{AP})(c_R^{AP} + 3c_P^{AP}) + (1 - p_R^{AP})(1 - p_E^{AP}) \cdot 4c_P^{AP}\Big],
\end{aligned}
$$

where c_j^k denotes the number of connections candidate i has with members of group $j \in \{R, E, P\}$ in pool $k \in \{FP, AP\}$.

Appendix C — Randomization and first stage results

This appendix reports diagnostic evidence supporting the identification strategy. We document balance under random committee assignment and present first-stage estimates for the individual- and department-level specifications.

Table C1: Balance under random committee assignment, individual level

Panel A.	Not controlling for expected connections			
	Papers	Books	Chapters	AIS
	(1)	(2)	(3)	(4)
Strong ties	0.004	-0.011	-0.022***	0.008
	(0.007)	(0.007)	(0.007)	(0.007)
Panel B.	**Controlling for expected connections**			
	Papers	Books	Chapters	AIS
	(1)	(2)	(3)	(4)
Strong ties	-0.000	-0.004	-0.000	0.000
	(0.010)	(0.010)	(0.010)	(0.011)

Notes: The table reports OLS estimates from regressions of candidate productivity measures on the number of strong ties between the candidate and committee members, using a sample of 31,243 applications to Associate and Full Professor positions. Panel A excludes controls for expected connections; Panel B includes them. The outcome variables are the number of journal articles, books, book chapters, and total AIS at the time of the exam, all standardized at the exam level to have mean zero and unit standard deviation. Standard errors are clustered at the exam level.
*** p < .01, ** p < .05, * p < .1

Table C2: Balance under random committee assignment, department level

Panel A.	Not controlling for expected connections				
	Candidates	Papers (z)	AIS (z)	Chapters (z)	Books (z)
Strong ties	0.504**	0.792***	0.638***	0.568***	0.493***
	(0.219)	(0.110)	(0.117)	(0.107)	(0.109)
Panel B.	**Controlling for expected connections**				
	Candidates	Papers (z)	AIS (z)	Chapters (z)	Books (z)
Strong ties	-0.313	-0.092	-0.096	-0.050	0.003
	(0.194)	(0.099)	(0.103)	(0.096)	(0.099)
Observations	13,215	13,215	13,215	13,215	13,215

Notes: The table reports OLS estimates at the department–exam level for applications to Associate and Full Professor positions. The dependent variables are the number of candidates in each department–exam cell and department-level standardised measures of productivity at the time of the exam. The main explanatory variable is the average number of strong ties between candidates and committee members. Panel A excludes controls for expected connections; Panel B includes them. All specifications control for the number of Associate and Full Professors in the department one year prior to the exam, as well as examination and year fixed effects. Standard errors are clustered at the department level. *** $p < .01$, ** $p < .05$, * $p < .1$

Table C3: Balance under random committee assignment, department level, by gender

Panel A.	Not controlling for expected connections					
	Female candidates	Male candidates	Papers (z)	AIS (z)	Chapters (z)	Books (z)
Female strong ties	4.268***	-3.124***	0.793***	0.616***	0.361**	0.317*
	(0.276)	(0.179)	(0.182)	(0.191)	(0.177)	(0.187)
Male strong ties	-1.733***	1.704***	0.503***	0.430***	0.477***	0.353***
	(0.098)	(0.179)	(0.119)	(0.130)	(0.121)	(0.121)
Joint F-test [p-value]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.005]
Panel B.	**Controlling for expected connections**					
	Female candidates	Male candidates	Papers (z)	AIS (z)	Chapters (z)	Books (z)
Female strong ties	0.052	0.045	-0.067	-0.020	-0.365*	-0.032
	(0.272)	(0.201)	(0.200)	(0.207)	(0.193)	(0.188)
Male strong ties	-0.132	-0.337*	-0.070	-0.111	0.123	0.045
	(0.084)	(0.174)	(0.126)	(0.137)	(0.127)	(0.131)
Joint F-test [p-value]	[0.293]	[0.153]	[0.777]	[0.695]	[0.140]	[0.937]
Observations	13,215	13,215	13,215	13,215	13,215	13,215

Notes: The table reports OLS estimates at the department–exam level for applications to Associate and Full Professor positions, separately by candidate gender. The dependent variables are the number of female and male candidates in each department–exam cell and department-level standardised measures of productivity at the time of the exam. The main explanatory variables are the average number of strong ties between female and male candidates and committee members. Panel A excludes controls for expected connections; Panel B includes them. All specifications control for the number of female and male Associate and Full Professors in the department one year prior to the exam, as well as examination and year fixed effects. The joint F-test tests whether both instruments are simultaneously zero. Standard errors are clustered at the department level.
*** p < .01, ** p < .05, * p < .1

Table C4: First-stage estimates, individual level

	Full Professor Exams		Associate Professor Exams	
Outcome: Qualified	(1)	(2)	(3)	(4)
Strong Connections	0.052***	0.052***	0.064***	0.064***
	(0.005)	(0.005)	(0.006)	(0.006)
Productivity controls:				
- Journal Articles		0.025***		0.037***
		(0.005)		(0.004)
- Total AIS		0.025***		0.008**
		(0.004)		(0.004)
- Book Chapters		0.014***		0.014***
		(0.003)		(0.003)
- Books		0.001		0.001
		(0.003)		(0.003)
Observations	13439	13439	17762	17762
KP rk Wald F-stat	120.147	126.545	129.455	131.548

Notes: The table reports first-stage estimates of Equation (1), where the outcome is an indicator for qualification and the explanatory variable is the number of strong ties to committee members. *Strong connections* is the number of strong ties (coauthorship, institutional affiliation, or doctoral supervision) to evaluators. All specifications control for the number of expected connections, exam fixed effects, and qualification-year fixed effects. Productivity variables are measured at the time of the exam and standardized within exam. The sample excludes 42 applications from eight candidates promoted in a previous wave who reapplied and were removed from the selection process. Standard errors are clustered at the exam level. The bottom row reports the Kleibergen–Paap rk Wald F-statistic for weak instruments.
*** p < .01, ** p < .05, * p < .1

Table C5: First-stage estimates, department level

	Qualified	Women Qualified	Men Qualified	Feminised Qualified	Not Feminised Qualified
	(1)	(2)	(3)	(4)	(5)
Panel A. Applications to Associate Professor positions					
Strong Connections	0.151***				
	(0.014)				
Female Strong Connections		0.135***	0.020*		
		(0.020)	(0.011)		
Male Strong Connections		0.004	0.143***		
		(0.006)	(0.018)		
Feminised Strong Connections				0.130***	0.004
				(0.017)	(0.006)
Not Feminised Strong Connections				0.031***	0.108***
				(0.010)	(0.019)
Observations	6732	6732	6732	6177	6177
KP rk Wald F-stat	116.501	34.303	34.303	18.573	18.573
KP rk LM statistic	103.181	67.370	67.370	33.519	33.519
SW F-stat	125.424	61.266	70.905	56.025	38.285
Panel B. Applications to Full Professor positions					
Strong Connections	0.113***				
	(0.011)				
Female Strong Connections		0.093***	0.017		
		(0.019)	(0.013)		
Male Strong Connections		0.001	0.114***		
		(0.003)	(0.012)		
Feminised Strong Connections				0.097***	0.013**
				(0.014)	(0.006)
Not Feminised Strong Connections				0.008	0.091***
				(0.006)	(0.016)
Observations	6483	6483	6483	6161	6161
KP rk Wald F-stat	108.433	13.781	13.781	23.844	23.844
KP rk LM statistic	103.774	26.542	26.542	45.758	45.758
SW F-stat	117.796	31.084	91.032	49.968	37.940

Notes: The unit of observation is a department–exam. Panel A includes applications to Associate Professor positions, while Panel B to Full Professor positions. Column (1) reports department-level first-stage estimates from Equation (3), where the outcome is the total number of qualified candidates. Columns (2) and (3) report first-stage estimates from Equation (5), where the outcomes are the number of qualified female and male candidates, respectively. Columns (4) and (5) report first-stage estimates from the feminised-equivalent version of Equation (5) described in Section 6.4. *Strong connections* measures the average number of connections between candidates in a given department–exam and committee evaluators, with female-, male-, feminised- and not-feminised-specific variants defined analogously for the respective subgroups. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level. The table reports F-statistics for weak and underidentification of the instrumental variables.
*** p < .01, ** p < .05, * p < .1

Appendix D — Alternative instrument definitions

This appendix describes alternative ways of measuring candidate–committee connections and documents how inference changes across instrument definitions. We define a connection between candidate i and evaluator c if, prior to the exam, (1) c supervised i's PhD thesis, (2) i and c were affiliated with the same university, or (3) i and c coauthored a publication. Comparing results across these definitions helps motivate the composite measure of tie intensity used in the main analysis.

D.1 Instrument variants

Binary exposure to a connected evaluator. We first construct an indicator equal to one if a candidate is connected to at least one committee member. Conditional on the expected probability of drawing a connected evaluator, this exposure is exogenous. We report the formula to compute this probability in Appendix B. The binary instrument yields a transparent LATE interpretation, but it collapses meaningful variation: among candidates with any connections, 27% are connected to more than one evaluator.

Count of strongly connected evaluators. To capture the intensity of exposure, we define the count of evaluators on the realized committee who are connected to candidate i in exam e:

$$\text{Strong ties}_{i,e} = \sum_{c \in e} \text{drawn}_{c,e} \times \text{connected}_{ic},$$

where $\text{drawn}_{c,e}$ indicates whether evaluator c is drawn to serve as committee member of exam e, and connected_{ic} indicates whether i and c share any connection. This variable is exogenous conditional on the expected share of connected evaluators. While it preserves variation in the number of connected evaluators, it does not materially improve precision relative to the binary instrument and still treats all connections as equally strong.

Count of connected evaluators separately by tie type. We next allow for heterogeneity in tie strength by including the count of connected evaluators separately for each connection type (shared affiliation, PhD supervision, and coauthorship). This specification delivers the most precise second-stage estimates. However, because it relies on multiple instruments, first-stage strength deteriorates, particularly in subgroup analyses.

Composite tie intensity (preferred). To balance interpretability, precision, and first-stage strength, we construct a composite measure of tie intensity:

$$\text{Strong tie intensity}_{i,e} = \sum_{j \in e} \text{drawn}_{j,e} \times (\text{same uni}_{ij} + \text{PhD advisor}_{ij} + \text{coauthors}_{ij}).$$

This measure retains variation in tie strength while avoiding the loss of power from estimating multiple first-stage coefficients. It is exogenous conditional on the expected total tie intensity; the formula reported in Section B extends equivalently to this case. Appendix C reports the balance of the randomization assignment.

D.2 Comparison across specifications

Appendix Tables D1 and D2 compare first- and second-stage estimates across instrument definitions. Across columns, the dependent variable is an indicator of whether a candidate applying for an Associate Professor position is promoted within three years of the national qualification exam; the endogenous regressor is the candidate's qualification status.

The specification with counts of connected evaluators, separately by tie type, delivers the most precise second-stage estimates but exhibits weaker first-stage performance, especially in split samples. The composite intensity measure achieves nearly equivalent second-stage precision while preserving a strong first stage. We therefore adopt the composite measure as our preferred specification.

Table D1: First-stage estimates, alternative instrument definitions

	Binary	Count	Separate	Intensity
Outcome: Qualified	(1)	(2)	(3)	(4)
At Least One Strong Tie	0.088***			
	(0.009)			
Strong Ties		0.063***		
		(0.006)		
PhD Advisors			0.124***	
			(0.023)	
Colleagues			0.048***	
			(0.006)	
Coauthors			0.095***	
			(0.021)	
Strong Tie Intensity				0.064***
				(0.006)
Observations	17762	17762	17762	17762
KP rk Wald F-stat	105.046	97.754	47.922	131.548

Notes: The table reports first-stage estimates of candidate qualification on connections with the evaluation committee, using four alternative instrument definitions. Column (1) includes an indicator for having at least one connected evaluator in the evaluation committee; column (2) includes the count of strongly connected evaluators; column (3) includes the count of connected evaluators, separately by tie type (shared affiliation, PhD supervision, and coauthorship); and column (4) considers a composite index equal to the number of tie types per evaluator, summed across evaluators. All specifications control for expected connections and candidate productivity (journal articles, total AIS, book chapters, and books), and include exam and result-year fixed effects. The sample consists of applications for Associate Professor positions. Standard errors are clustered at the exam level.

*** $p < .01$, ** $p < .05$, * $p < .1$

Table D2: Second-stage estimates, alternative instrument definitions

	Binary	Count	Separate	Intensity
Outcome: Promotion within 3 yrs	(1)	(2)	(3)	(4)
Qualified Candidate	0.515***	0.603***	0.586***	0.588***
	(0.116)	(0.121)	(0.089)	(0.096)
Observations	17762	17762	17762	17762

Notes: The table reports 2SLS estimates using the alternative instrument definitions described in Table D1. The endogenous regressor is candidate qualification and the outcome is an indicator for promotion within three years of the exam. All specifications control for expected connections, candidate productivity, and include exam and result-year fixed effects. Standard errors are clustered at the exam level.

*** $p < .01$, ** $p < .05$, * $p < .1$

Appendix E — Compliers

This appendix characterizes the complier population induced by random committee assignment. We document how connected and unconnected candidates differ among qualifiers and non-qualifiers, estimate the shares and observable characteristics of always-takers, compliers, and never-takers, and compare OLS and 2SLS estimates to clarify interpretation.

Connected vs. unconnected candidates. Appendix Table E1 summarizes observable differences between candidates with and without committee connections, separately for the full sample and by qualification outcome. Two patterns stand out. First, candidates who qualify, whether connected or not, are substantially more productive than the average applicant. Second, among qualifiers, connected candidates are on average less productive than unconnected ones, consistent with connections affecting the outcomes of marginal but still highly productive candidates.

Always-takers, compliers, and never-takers. Following Imbens and Rubin, 1997, we characterize the complier, always-taker, and never-taker subpopulations in our IV framework, adapting their approach to our setting in which the instrument is exogenous conditional on a control variable.

For ease of interpretation, we use a binary instrument $Z_i \in \{0, 1\}$ indicating whether candidate i has a connection on the committee, which is exogenous conditional on $P_i \in [0, 1]$, the probability of drawing a connected evaluator. We use this instrument to isolate exogenous variation in qualification status $Q_i \in \{0, 1\}$.

To build intuition, suppose first that Z were unconditionally random. With a binary instrument and binary treatment, and assuming monotonicity (no defiers), Imbens and Rubin, 1997 show that the shares of the principal strata are identified from observable cells. Candidates who receive the instrument but do not qualify must be never-takers, while those who qualify without receiving the instrument must be always-takers. The corresponding population shares are therefore:

$$\phi_a = \mathbb{P}(Q = 1 \mid Z = 0), \qquad \phi_n = \mathbb{P}(Q = 0 \mid Z = 1), \qquad \phi_c = 1 - \phi_a - \phi_n.$$

Similarly, the distributions of an observable characteristic X for always-takers and never-takers are given by

$$F_a(X) = F(X \mid Q = 1, Z = 0), \qquad F_n(X) = F(X \mid Q = 0, Z = 1).$$

By subtraction, the distribution of X for compliers can be recovered as

$$F_c(X) = \frac{1}{\phi_c} F(X) - \frac{\phi_a}{\phi_c} F_a(X) - \frac{\phi_n}{\phi_c} F_n(X).$$

In our setting, identification relies on conditional randomization: among candidates with a similar probability of drawing a connected evaluator, some randomly realize such a connection while others do not. We therefore estimate subpopulation shares and the distribution of a pre-exam observable X_i within bins of P_i (Appendix Figure E1), and then aggregate these estimates across bins using the empirical distribution of P_i to recover population-level distribution in the entire sample.

To implement this approach under conditional randomization, we approximate conditioning on P_i by discretizing the support of P into 50 bins, each containing roughly 700 candidates with a similar probability of drawing a connected evaluator.[23] Let $b \in B$ index these bins. Because the instrument is randomly assigned conditional on P_i, it is approximately random within each bin.

Within each bin we compute the shares of the principal strata using the same logic as in the

23. The first bin contains all candidates with a probability equal to zero and is substantially larger, with nearly 4,000 observations.

unconditional case:

$$\phi_a^b = \mathbb{P}(Q = 1 \mid Z = 0, B = b), \qquad \phi_n^b = \mathbb{P}(Q = 0 \mid Z = 1, B = b), \qquad \phi_c^b = 1 - \phi_a^b - \phi_n^b.$$

Weighting by the number of candidates in each bin, N^b, we recover the population-level share of compliers by integrating over the empirical distribution of P_i:

$$\phi_c = \frac{1}{N} \sum_{b \in B} N^b \phi_c^b,$$

where $N = \sum_{b \in B} N^b$.

Similarly, we recover the distribution of pre-exam characteristics for each principal stratum by aggregating the bin-specific distributions. Let $F_c^b(X)$ denote the distribution of X among compliers in bin b, computed using the Imbens and Rubin, 1997 decomposition within the bin. The population distribution of X for compliers is then

$$F_c(X) = \frac{1}{N\phi_c} \sum_{b \in B} N^b \phi_c^b F_c^b(X).$$

This approach approximates conditioning on the continuous variable P_i by discretizing its support. As the partition becomes finer, the resulting estimators converge to the population quantities implied by conditioning on P_i. In practice, our results are robust to alternative bin definitions and numbers of bins.

Appendix Table E2 reports the estimated shares of the principal strata. The vast majority of candidates (82.5%) are never-takers who do not qualify regardless of whether they draw a connected evaluator, reflecting the highly competitive nature of the *habilitación* system. Always-takers and compliers account for roughly 9% and 8.5% of the sample, respectively. The estimated complier share is consistent with the magnitude of the first-stage effect (Appendix Table D1, Column 1).

Always-takers are approximately 0.5 standard deviations more productive than the average candidate, whereas compliers are about 0.1 standard deviations more productive than average and tend to come from smaller departments that experienced slower growth in the preceding five years. Consistent with the tournament nature of the qualification system, never-takers exhibit the lowest productivity on average among the three subpopulations.

Appendix Figure E2 shows how the three subpopulations are distributed along the research output distribution, where publications are normalized within each applicant pool. We estimate these densities using the same mixture logic as above, discretizing both the probability of drawing a strong tie and the distribution of X, and computing the mass of each subpopulation within each cell. Candidates in cells with insufficient mass to identify subpopulation shares are classified as unidentifiable.[24]

The figure reveals a clear ordering across compliance types. Always-takers first-order stochastically dominate both compliers and never-takers, with substantially more mass in the upper tail of the productivity distribution. The distribution of compliers lies between the two extremes and exhibits near first-order stochastic dominance over never-takers, with only a small violation in the far left tail.

Appendix Figure E3 reports the same densities separately by gender. Although the smaller number of female candidates leads to a larger share of unidentifiable observations, the overall patterns are similar. Never-takers are disproportionately concentrated among lower-productivity applicants, whereas always-takers and compliers account for a larger share of candidates with above-average research output.

As an alternative ranking, we estimate a logit model predicting qualification status using a broader set of observables – including total papers, books, book chapters, Web of Science publications, and measures of connectedness – and use the predicted index to rank candidates within each exam:

$$\mathbb{P}[Q_{i,e} = 1 \mid X_{i,e}] = \Lambda(X_{i,e}\beta), \tag{E.1}$$

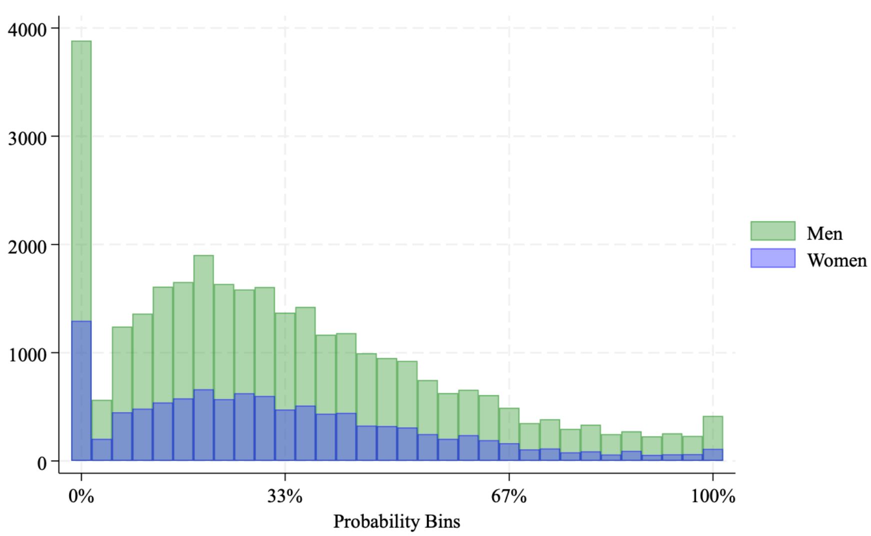
where Λ denotes the logistic function.

24. There is a trade-off involved: larger cells reduce the risk of insufficient mass but may group candidates with very different probabilities of drawing a strong tie, potentially weakening the conditional exogeneity assumption. In this analysis we prioritize the internal validity of the estimates and tolerate a larger share of unidentifiable candidates.

Appendix Figure E4 plots the distribution of subpopulations along this predicted ranking. Always-takers and compliers are concentrated among higher-ranked candidates, whereas never-takers are disproportionately represented toward lower ranks. Above the qualification margin, always-takers exhibit the highest density, followed by compliers. These patterns are similar for both men and women (Appendix Figure E5).

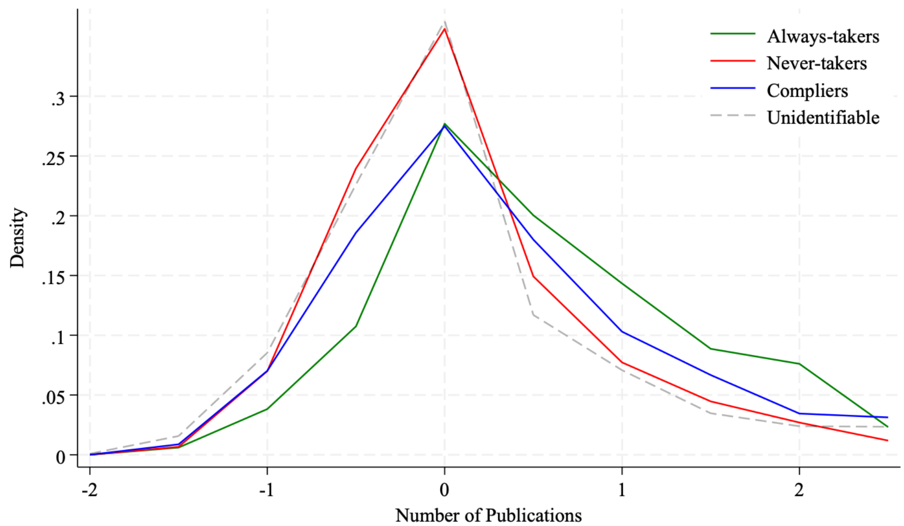
OLS vs. 2SLS. Appendix Table E3 reports OLS and 2SLS estimates for promotion outcomes. For Associate Professor candidates, the two estimates are similar in magnitude, but this should not be interpreted as evidence of limited selection. OLS is upward biased because qualification is correlated with unobserved strength and long-run promotion potential. In contrast, 2SLS identifies a local average treatment effect for marginal candidates whose careers are particularly sensitive to the qualification bottleneck. The proximity of the estimates therefore likely reflect offsetting forces rather than the absence of selection.

Figure E1: Distribution of the probability of drawing a connected evaluator



Notes: The figure shows the distribution of male and female candidates by their probability of drawing at least one strong tie on the evaluation committee. Strong ties include coauthors, departmental colleagues, and PhD supervisors.

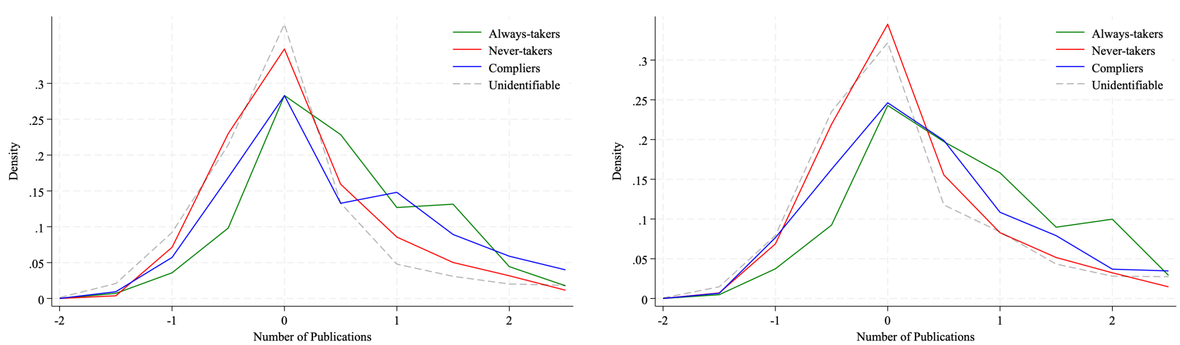
Figure E2: Productivity distribution by subpopulation



Notes: The figure shows the distribution of instrumental-variable subpopulations – compliers, always-takers, and never-takers – over candidates' publication counts in the year prior to the exam. Publication counts are normalized within exam. The fourth category, *unidentifiable*, denotes candidates in cells with too few observations to identify the shares of IV subpopulations.

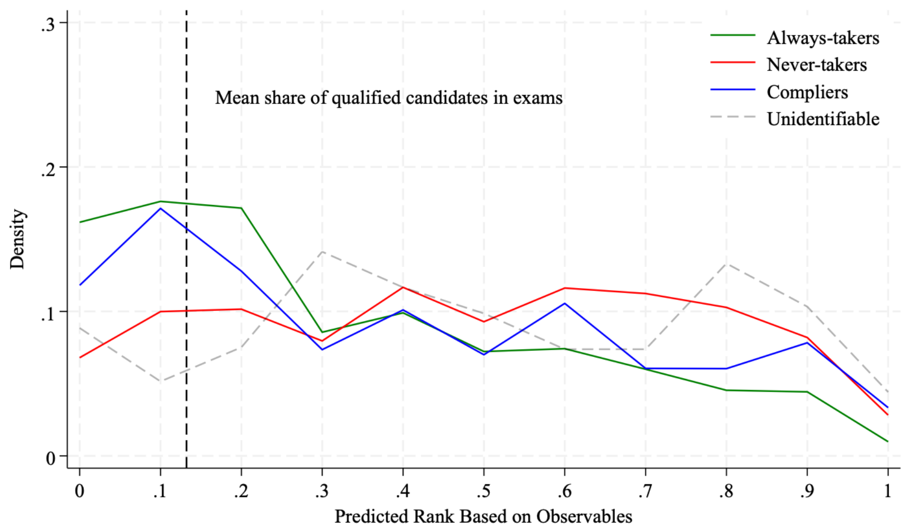
Figure E3: Productivity distribution by subpopulation and gender

A) Women B) Men



Notes: The figure shows the distribution of instrumental-variable subpopulations – compliers, always-takers, and never-takers – over candidates' publication counts in the year prior to the exam, separately by gender. Panel A reports the distribution for female candidates, and Panel B reports the distribution for male candidates. Publication counts are normalized within exam. The fourth category, *unidentifiable*, denotes candidates in cells with too few observations to identify the shares of IV subpopulations.

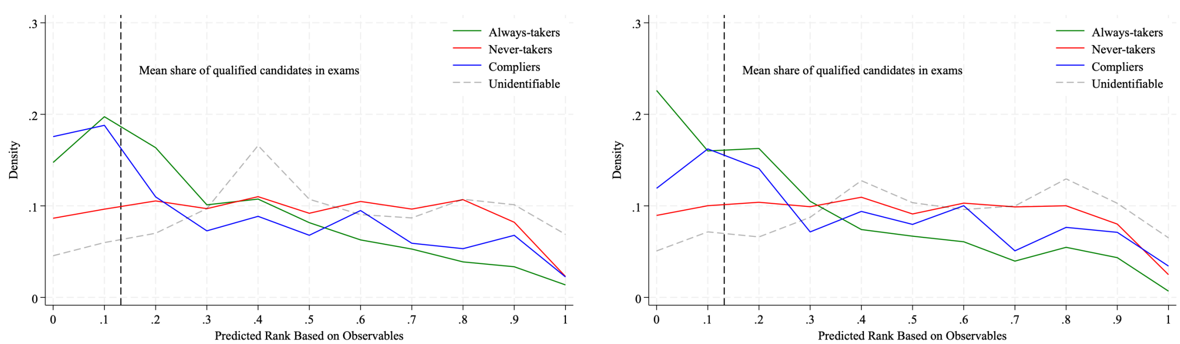
Figure E4: Predicted qualification rank distribution by subpopulation



Notes: The figure shows the distribution of instrumental-variable subpopulations – compliers, always-takers, and never-takers – over candidates' predicted rank based on the model in Equation (E.1). Candidates are ranked within each exam using the predicted qualification probability from the logit model. The fourth category, *unidentifiable*, denotes candidates in cells with too few observations to identify the shares of IV subpopulations.

Figure E5: Predicted qualification rank distribution by subpopulation and gender

A) Women B) Men



Notes: The figure shows the distribution of instrumental-variable subpopulations –compliers, always-takers, and never-takers – over candidates' predicted rank based on the model in Equation (E.1), separately by gender. Candidates are ranked within each exam using the predicted qualification probability from the logit model. Panel A reports the distribution for female candidates, and Panel B reports the distribution for male candidates. The fourth category, *unidentifiable*, denotes candidates in cells with too few observations to identify the shares of IV subpopulations.

Table E1: Candidate characteristics by committee connections and qualification outcome

	Connected (1)	Not Connected (2)	Connected - Not Connected (3)	SE (4)
Panel A: All Candidates				
Number of observations:	9911	21332		
Papers	0.01	-0.00	0.02	0.01
Papers in Web of Science	0.01	-0.00	0.01	0.01
Books	-0.00	0.00	-0.00	0.01
Chapters	-0.01	0.01	-0.02*	0.01
Total AIS	0.02	-0.01	0.02**	0.01
Average AIS	0.01	-0.01	0.02	0.01
Citations	0.02	-0.01	0.04***	0.01
Panel B: Successful Candidates				
Number of observations:	1744	1829		
Papers	0.31	0.38	-0.07**	0.04
Papers in Web of Science	0.26	0.33	-0.08**	0.04
Books	0.09	0.06	0.03	0.03
Chapters	0.17	0.20	-0.03	0.03
Total AIS	0.23	0.31	-0.08**	0.04
Average AIS	0.12	0.20	-0.08**	0.03
Citations	0.25	0.34	-0.09**	0.04
Panel C: Unsuccessful Candidates				
Number of observations:	8167	19503		
Papers	-0.05	-0.04	-0.01	0.01
Papers in Web of Science	-0.05	-0.03	-0.01	0.01
Books	-0.02	-0.01	-0.01	0.01
Chapters	-0.05	-0.01	-0.04***	0.01
Total AIS	-0.03	-0.04	0.01	0.01
Average AIS	-0.01	-0.03	0.01	0.01
Citations	-0.02	-0.04	0.02*	0.01

Notes: The unit of observation is an application to the *habilitación* qualification system. A candidate is classified as connected if at least one committee member has a strong tie to the candidate (coauthorship, shared departmental affiliation, or PhD supervision). Panel A includes all applicants; Panel B includes those who qualified; Panel C includes those who did not. Columns (1) and (2) report mean productivity measures standardized within exam; column (3) reports the difference between connected and unconnected candidates and column (4) the corresponding standard error.
*** $p < .01$, ** $p < .05$, * $p < .1$

Table E2: Characteristics of compliers, always-takers, and never-takers

	Share sub-population	Number of Papers	Article Influence Score	Department Size	Department Growth
Compliers	8.4 %	0.13	0.04	12.39	1.68
Always-takers	9.1 %	0.41	0.35	15.35	3.58
Never-takers	82.4 %	-0.04	-0.04	17.28	5.21

Notes: The table reports the share and mean characteristics of each subpopulation, estimated following Imbens and Rubin, 1997 as described in Appendix E. Number of papers and Article Influence Score are standardized within exam to have mean zero and unit standard deviation. Department size is the number of tenured faculty one year prior to the exam. Department growth is the change in the number of tenured faculty over the five years prior to the exam.

Table E3: Individual-level estimates, OLS and 2SLS

	Applicants to Associate Professor		Applicants to Full Professor	
	(1)	(2)	(3)	(4)
	OLS	2SLS	OLS	2SLS
Qualified	0.329***	0.400***	0.269***	0.057
	(0.009)	(0.110)	(0.008)	(0.111)
Observations	17762	17762	13439	13439
R^2	0.14	0.07	0.17	0.06
KP rk Wald F-stat		131.548		126.545

Notes: The table reports OLS and 2SLS estimates of the effect of qualification on the probability of promotion within 15 years of the exam, using a sample of 31,201 applications to Associate and Full Professor positions. All specifications control for standardized productivity measures at the time of the exam and expected connections. Standard errors are clustered at the exam level.
*** p < .01, ** p < .05, * p < .1

Appendix F — Further results

This appendix reports additional results that complement the main analysis presented in the paper.

First, we present additional individual-level estimates of the impact of qualification on subsequent promotion probabilities. Second, we report supplementary department-level evidence illustrating the contrasting effects of qualification shocks at the Associate Professor and Full Professor levels. Third, we provide additional results describing how female qualifications to Associate Professor affect departmental faculty composition and promotion dynamics. Finally, we present heterogeneity analyses by field group (STEM+M versus Social Sciences and Humanities) and using an alternative exposure measure based on research-topic feminization.

F.1 Aggregate department-level effects of qualification shocks

Appendix Figure F1 summarizes the aggregate department-level effects of qualification shocks. Panel A shows that exogenous qualifications to Associate Professor generate persistent increases in cumulative promotions to the same rank. Panel B reports the corresponding effect for qualifications to Full Professor. In contrast, these generate only a short-run increase in Full Professor promotions.

F.2 Impact of female Associate Professor qualifications on promotions of 'non-candidates'

Appendix Figure F2 examines whether the department-level effects of female qualification to Associate Professor reported in Panel A of Figure 3 in the main text are mechanically driven by the promotion of the qualifying candidate herself. The outcome is cumulative department-level promotions *excluding* promotions of the department's own candidates in a given exam. While this exclusion reduces precision, the point estimates continue to indicate a potentially sizable trickle-down effect. No analogous effect is observed for men.

F.3 Impact of female Associate Professor qualifications on tenured female faculty

Appendix Figure F3 examines how female qualification shocks affect the gender composition of tenured faculty. Panel A shows that qualifying a woman to Associate Professor generates a sustained increase in the number of tenured female professors in the department. In contrast, Panel B shows no statistically significant effect following the exogenous qualification of a woman to Full Professor.

F.4 Impact of Full Professor qualifications on promotions to Full Professor

Appendix Figure F4 reports the event-study estimates of the impact of Full Professor qualification shocks on subsequent promotions to Full Professor within the department. There is a short-run increase in promotions immediately following the qualification exam, but it dissipates within a few years, and the estimates remain close to zero thereafter.

F.5 Impact of Full Professor qualifications on promotions to Associate Professor

Appendix Figure F5 reports event-study estimates of the effect of exogenous Full Professor qualifications on subsequent promotions to Associate Professor within the department. The four panels distinguish the gender of the qualifying candidate and the gender of promoted faculty. Across all panels, the estimates are statistically imprecise. We therefore find no evidence that qualifications to Full Professor systematically affect promotion dynamics at the Associate Professor level.

F.6 Heterogeneity of the impact of female qualifications by field group: STEM+M versus SSH.

The main text reports estimates of heterogeneous effects by field group ten years after the qualification exam. Appendix Figure F6 complements that analysis by showing the full event-study path separately for STEM+M and Social Sciences and Humanities (SSH).

F.7 Male outcomes from qualifications in female-intensive research areas

Appendix Figure F7 complements the main analysis of female-intensive qualifications by documenting the corresponding effects on male outcomes. The four panels show the impact of exogenously qualifying a candidate working in female-intensive research areas to *Associate Professor* on male promotions to Associate Professor, male PhD graduates, male research-active PhDs, and male PhDs subsequently promoted to Associate Professor. Across all outcomes, the effects on men are small and statistically insignificant throughout the fifteen-year horizon.

F.8 Effects of qualifications in non-feminised research areas

Appendix Figure F8 reports the department-level effects of exogenously qualifying a candidate working in a non-feminised research area to *Associate Professor*. The four panels show effects on female and male promotions to Associate Professor and on female and male PhD graduates. Consistent with the hypothesis that research-topic feminisation drives part of the trickle-down effect, qualifying researchers in non-feminised fields generates no discernible impact on female or male outcomes.

F.9 Horse race regression: faculty promotions

Appendix Figure F9 reports the results of a horse race specification that simultaneously includes both the number of qualified female candidates and the number of qualified candidates working in female-intensive research areas as endogenous regressors. The four panels show effects on cumulative female and male promotions to *Associate Professor*. The coefficient on female-intensive qualifications (Panel B) is positive and statistically significant, while the coefficient on female qualifications (Panel A) remains close to one but is less precisely estimated.

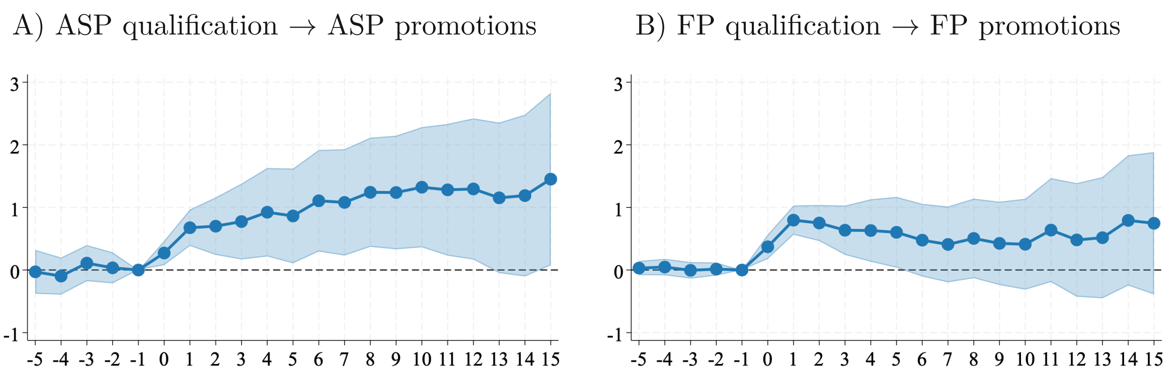
F.10 Horse race regression: PhD graduates

Appendix Figure F10 reports the horse race estimates for PhD outcomes, mirroring the specification in the preceding subsection. The four panels show effects on cumulative female and male PhD graduates. The effect of female qualifications on female PhD graduates (Panel A) remains large and stable even after controlling for the feminisation of research areas, while the coefficient on female-intensive qualifications (Panel B) is small and imprecise. The analogous effects on male PhD graduates are negligible across both channels.

F.11 Impact of female qualification to Associate Professor on PhD graduates

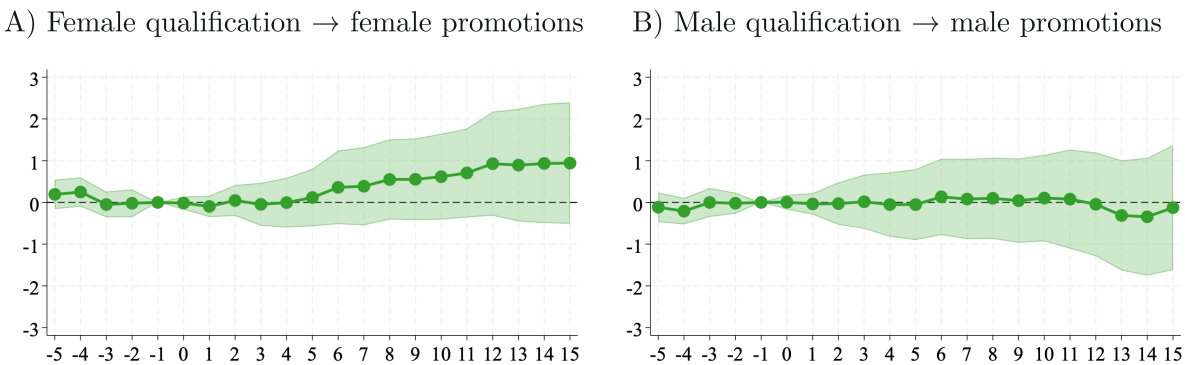
The main text documents that exogenous female qualifications lead to an increase in the number of female PhD graduates. Appendix Figure F11 complements this result by decomposing the effect according to whether the lead supervisor of the graduating student is a woman or a man.

Figure F1: Impact of exogenous qualifications on the department



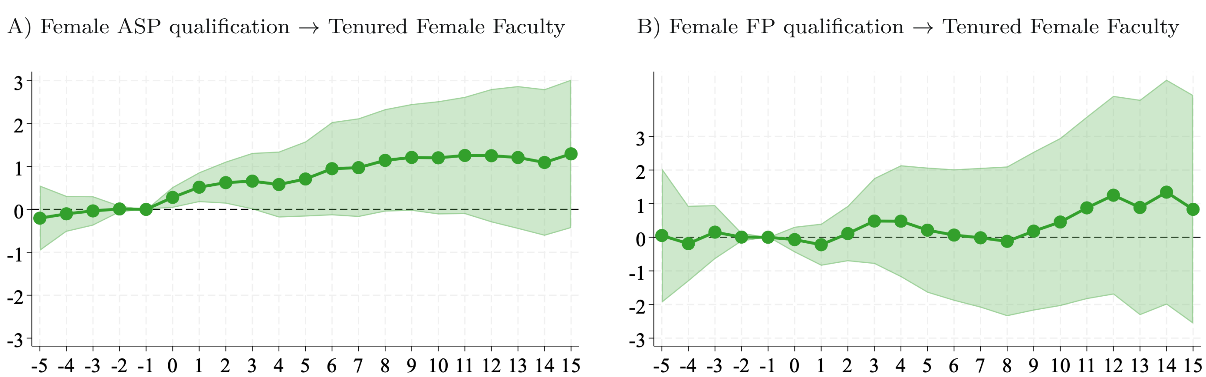
Notes: The figure shows 2SLS estimates of coefficient δ_1 from Equation 4, capturing the impact of exogenously qualifying an Associate Professor on the cumulative number of promotions to Associate Professor (Panel A), and the impact of exogenously qualifying a Full Professor on the cumulative number of promotions to Full Professor (Panel B), over the years following the qualification exam. The unit of analysis is department–exam. Standard errors clustered at the department level.

Figure F2: Department-level effects of female and male qualifications to *Associate Professor* on promotions to *Associate Professor*, excluding own promotions



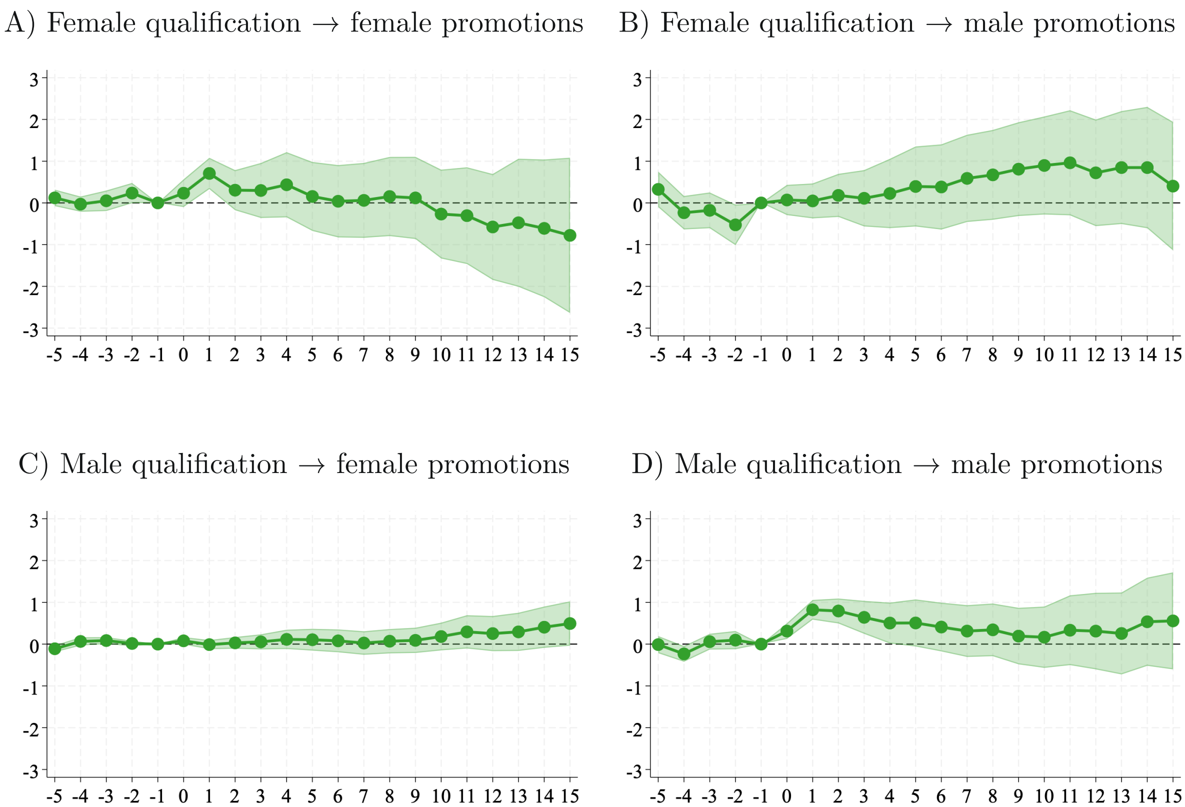
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6). Panel A shows the effect of an exogenous female qualification to *Associate Professor* on the cumulative number of female promotions to *Associate Professor*, while Panel B shows the impact of an exogenous male qualification on male promotions to *Associate Professor*. Outcomes are measured cumulatively over the years following the qualification exam. Promotion counts exclude promotions of departments' own candidates. The unit of analysis is the department–exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level.

Figure F3: Department-level effects of a female qualification on tenured female faculty



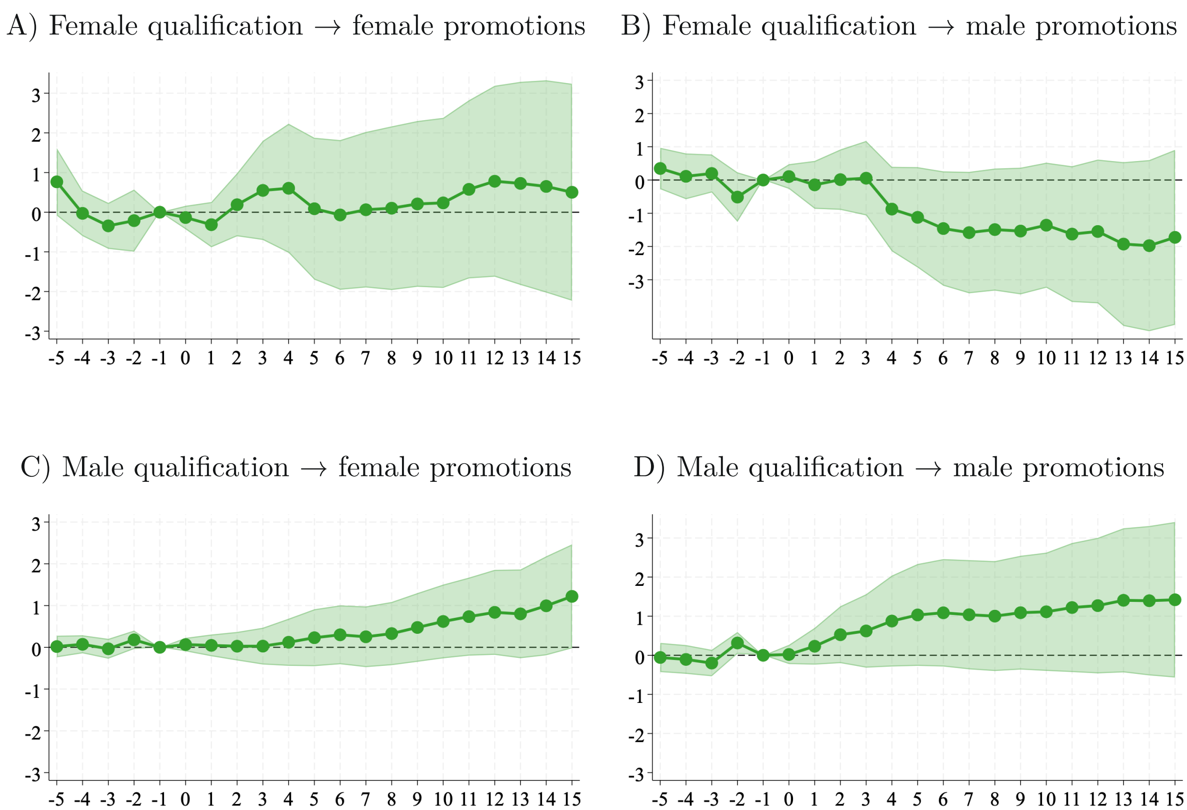
Notes: The figure reports 2SLS estimates of coefficient δ_1 from Equation (6), capturing the impact of exogenously qualifying a woman on departmental faculty outcomes. Panel A shows the effect of qualifying a woman to Associate Professor on the stock of tenured female faculty. Panel B shows the effect of qualifying a woman to Full Professor on the stock of tenured female faculty. Outcomes are measured cumulatively over the years following the qualification exam. The unit of analysis is the department–exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level.

Figure F4: Department-level effects of female and male qualifications to *Full Professor* on promotions to *Full Professor*



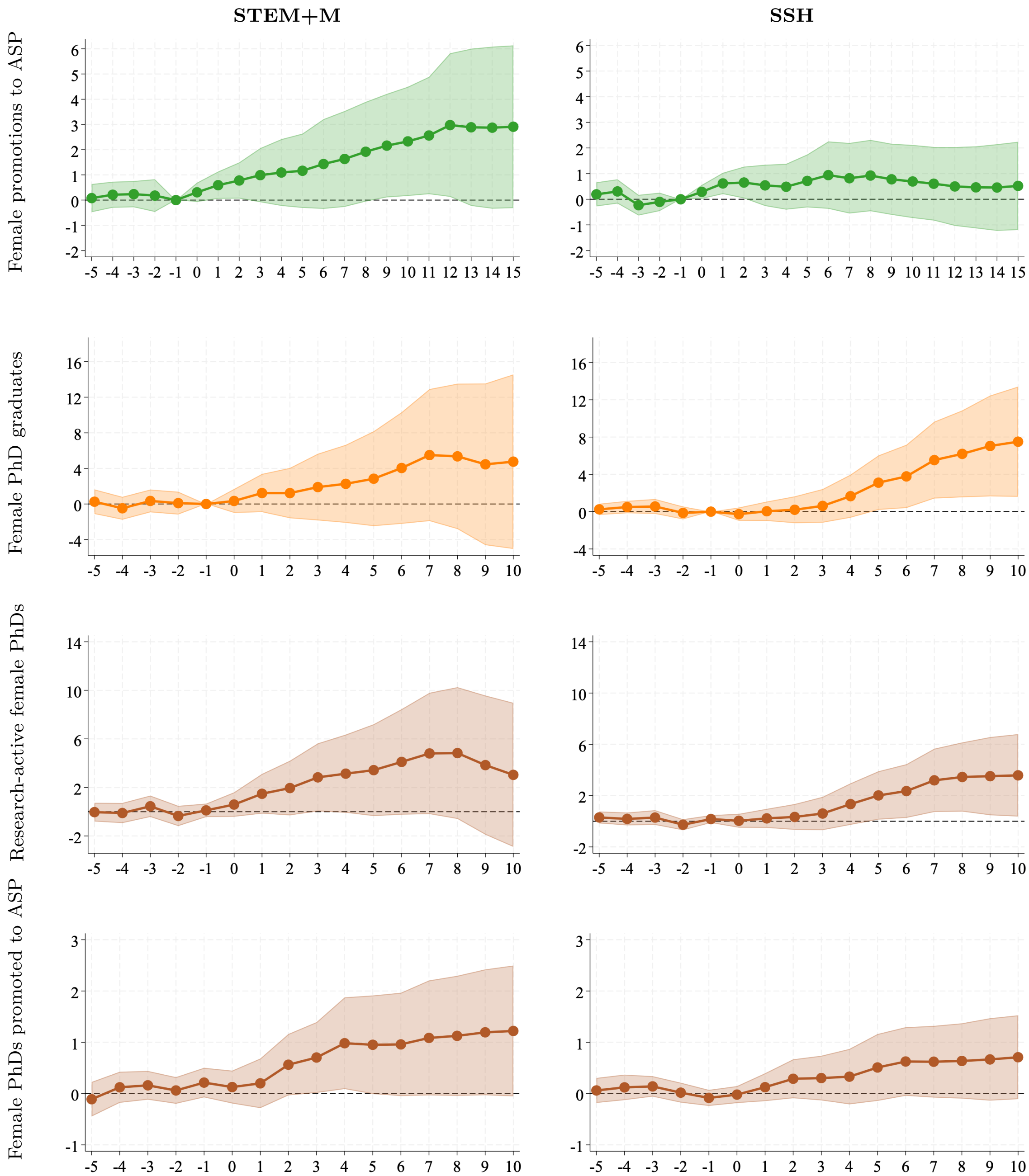
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6). Panels A and B show the effect of an exogenous female qualification to *Full Professor* on the cumulative number of female and male promotions to *Full Professor*, respectively. Panels C and D show the corresponding effects of an exogenous male qualification. Outcomes are measured cumulatively over the years following the qualification exam. The unit of analysis is the department–exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level.

Figure F5: Department-level effects of female and male qualifications to *Full Professor* on promotions to *Associate Professor*



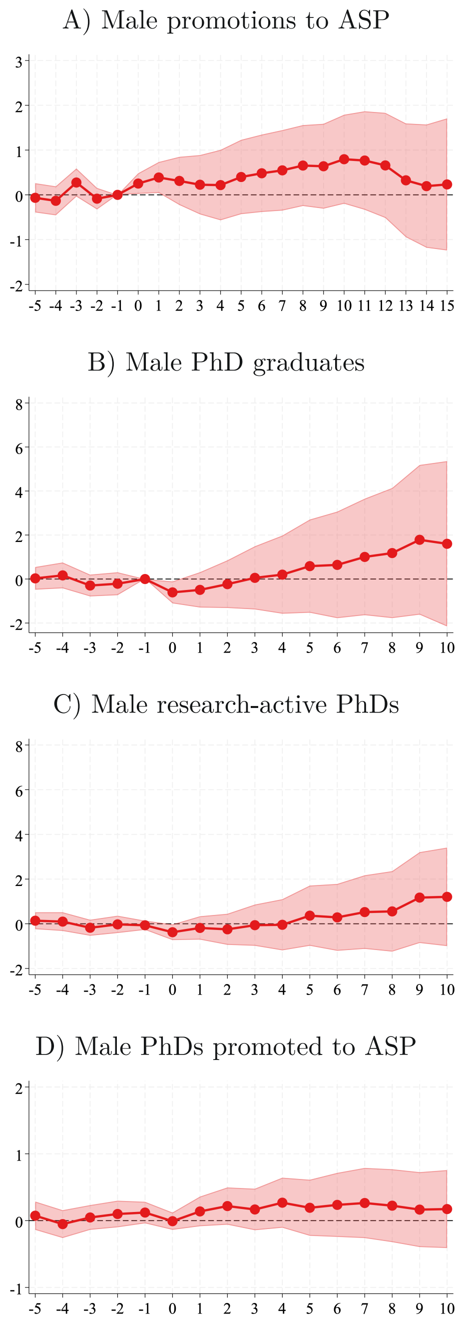
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6). Panels A and B show the effect of an exogenous female qualification to *Full Professor* on the cumulative number of female and male promotions to *Associate Professor*, respectively. Panels C and D show the corresponding effects of an exogenous male qualification. Outcomes are measured cumulatively over the years following the qualification exam. The unit of analysis is the department–exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level.

Figure F6: Department-level effects of female qualifications to *Associate Professor*: Heterogeneity by field group



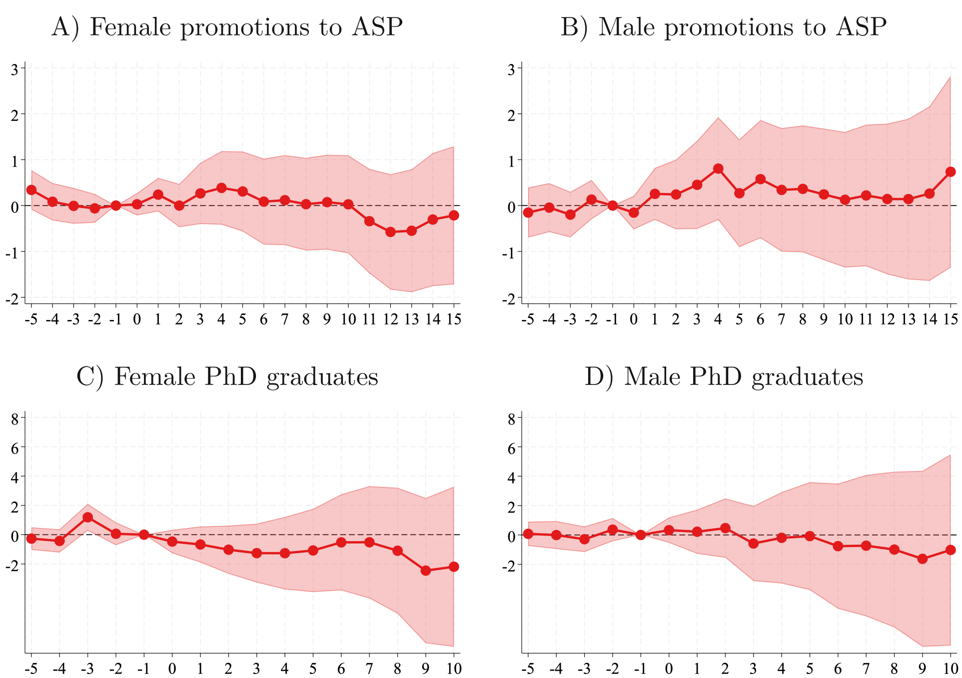
Notes: The figure reports 2SLS estimates of coefficient δ_1 from Equation 6, capturing the impact of exogenously qualifying a woman to *Associate Professor*. Rows 1–4 report effects on cumulative (i) promotions to Associate Professor, (ii) PhD graduates, (iii) PhD graduates remaining research-active five years after graduation, and (iv) PhD graduates promoted to Associate Professor within ten years. The left column reports estimates for STEM+M fields; the right column reports estimates for Social Sciences and Humanities (SSH). Outcomes are measured relative to the year of qualification. The unit of observation is the department–exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level.

Figure F7: Department-level effects of *ASP* qualifications in female-intensive research areas



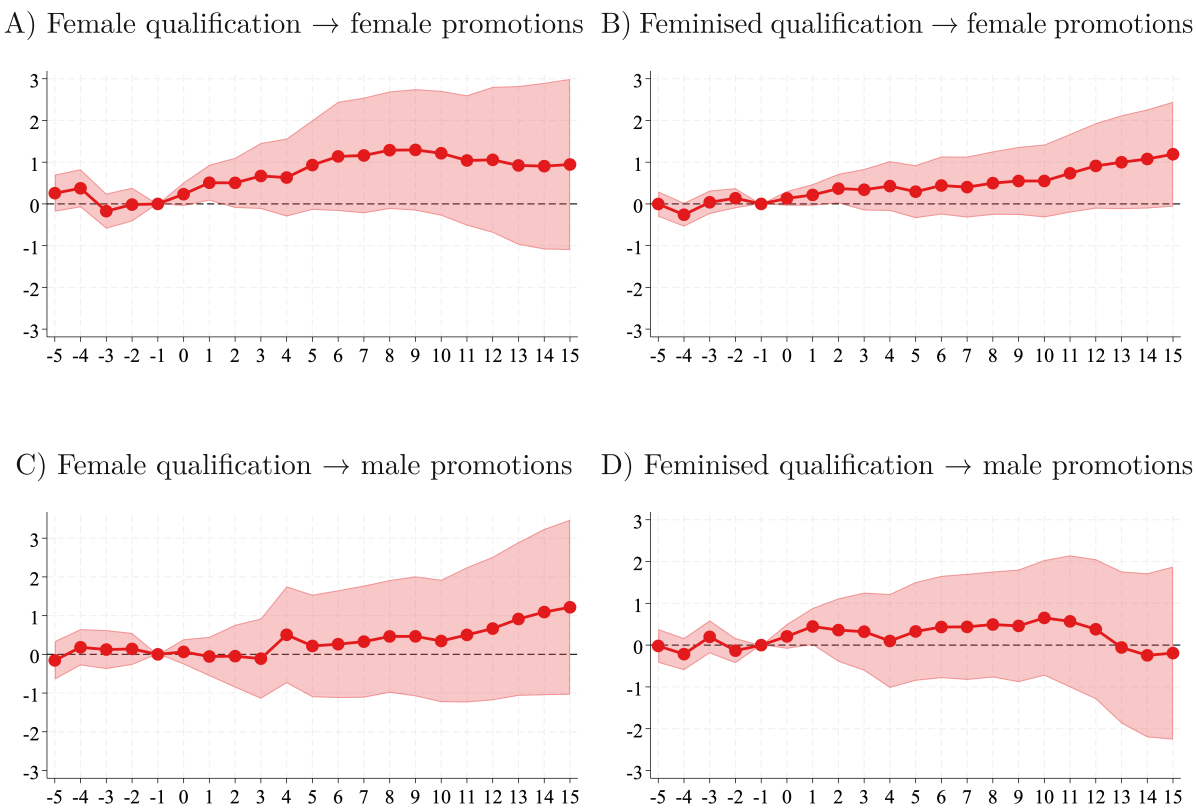
Notes: The figure reports 2SLS estimates of coefficient δ_1 from a specification analogous to Equation (6), where the endogenous variables are the number of qualified candidates working in female-intensive research areas and the number of those who do not. The coefficient captures the impact of exogenously qualifying such an applicant to *Associate Professor*. Each panel reports the dynamic effect of exogenously qualifying candidates working in more female-intensive research areas to *Associate Professor* on the outcome indicated in the panel title. Outcomes are measured relative to the year of qualification and are cumulative over time. The unit of observation is the department-exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure F8: Department-level effects of ASP qualifications of non-feminised candidates



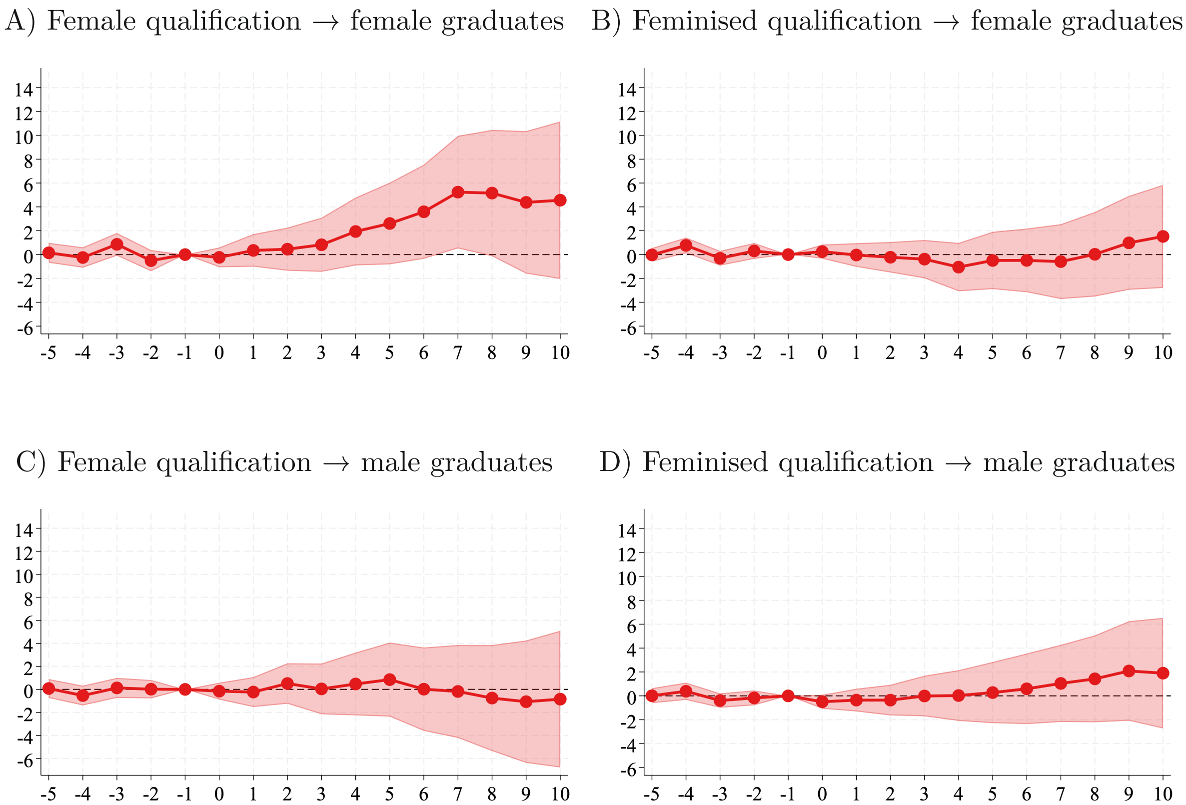
Notes: The figure reports 2SLS estimates of coefficient δ_2 from a specification analogous to Equation (6), where the endogenous variables are the number of qualified candidates working in female-intensive research areas and the number of those who do not. The coefficient captures the impact of exogenously qualifying such an applicant to *Associate Professor*. Each panel reports the dynamic effect of exogenously qualifying non-feminised candidates to *Associate Professor* on the outcome indicated in the panel title. Outcomes are measured relative to the year of qualification and are cumulative over time. The unit of observation is the department-exam cell. All specifications include controls listed in Section 5.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure F9: Horse race regression, including own promotions



Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from a specification analogous to Equation (6). The endogenous regressors are (i) the number of qualified female candidates and (ii) the number of qualified candidates in female-intensive research areas, allowing us to assess the relative roles of gender and research orientation. Panels A–B use as the outcome the cumulative number of female promotions to *Associate Professor*, measured relative to the year of the qualification exam. Panel A shows the effect of exogenously qualifying a woman; Panel B shows the effect of exogenously qualifying a researcher in a female-intensive field. Panels C–D report analogous estimates where the outcome is cumulative male promotions. In all panels, promotion counts include promotions of the department's own candidates. All specifications include controls listed in Section 5.2. The unit of observation is the department–exam cell, and standard errors are clustered at the department level.

Figure F10: Horse race regression

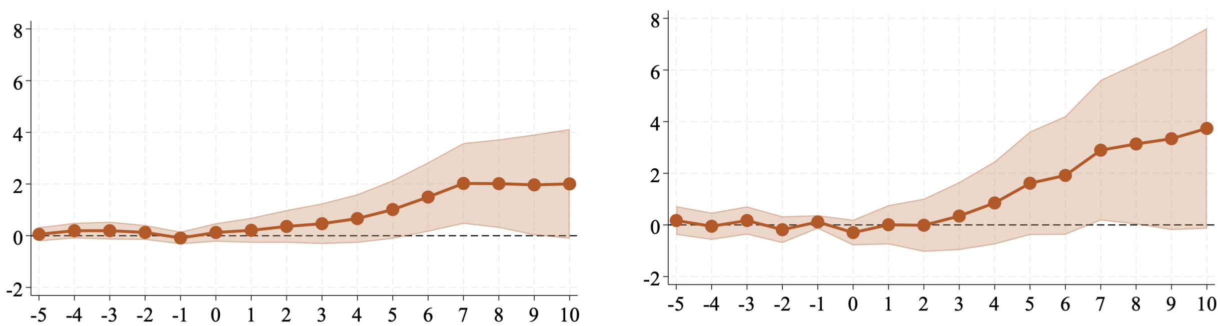


Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from a specification analogous to Equation (6). The endogenous regressors are (i) the number of qualified female candidates and (ii) the number of qualified candidates in female-intensive research areas, allowing us to assess the relative roles of gender and research orientation. Panels A–B use as the outcome the cumulative number of female PhD graduates, measured relative to the year of the qualification exam. Panel A shows the effect of exogenously qualifying a woman; Panel B shows the effect of exogenously qualifying a researcher in a female-intensive field. Panels C–D report analogous estimates where the outcome is cumulative male PhD graduates. All specifications include controls listed in Section 5.2. The unit of observation is the department–exam cell, and standard errors are clustered at the department level.

Figure F11: Department-level effects of female qualifications to *Associate Professor* on PhD graduates, by supervisor gender

A) Female qualification → Female PhDs with Female Lead Supervisor

B) Female qualification → Female PhDs with Male Lead Supervisor



Appendix G — A Simple Decomposition of the Trickle-Down Effect

The reduced-form estimates presented in the main text document substantial downstream effects of female promotions at the department level. In particular, the promotion of a woman to Associate Professor generates on average roughly one additional female promotion within the department over the subsequent fifteen years. While these estimates establish the presence of a trickle-down effect, they combine several distinct dynamic components: the direct promotion of the qualifying candidate, the delayed promotion of initially unsuccessful candidates who eventually catch up, and potential spillovers operating through the departmental environment.

This appendix provides a simple exercise that combines the individual-level promotion dynamics with the department-level reduced-form estimates in order to recover a stylized representation of the underlying spillover process.

G.1 Individual promotion dynamics

Let Δ_h denote the cumulative causal effect of qualification on the probability that the focal candidate has been promoted by horizon h. These estimates come from the individual-level analysis described in Section 6.

Define the annual flow of the treatment effect as

$$d_h = \Delta_h - \Delta_{h-1}, \qquad \Delta_{-1} = 0.$$

The quantity d_h measures the additional probability that the focal candidate is promoted in year h as a consequence of receiving the qualification shock. Because candidates who initially fail to qualify may later be promoted through other channels, d_h can become negative at longer horizons, reflecting catch-up promotions among initially unsuccessful candidates.

G.2 Department-level reduced form

Let Y_h denote the cumulative department-level effect of a female qualification shock on the number of female promotions within the department by horizon h. These estimates correspond to the department-level event-study coefficients reported in Section 6.

Define the annual department-level flow

$$y_h = Y_h - Y_{h-1}, \qquad Y_{-1} = 0.$$

The difference between y_h and d_h captures the additional promotions that occur within the department beyond the direct promotion of the qualifying candidate.

G.3 A simple spillover model

We assume that each additional female promotion generates a sequence of downstream effects on future female promotions within the department. Let b_j denote the number of additional female promotions generated in year j by the promotion of one additional woman today.

Under this representation, the annual department-level effect can be written as

$$y_h = d_h + \sum_{s=0}^{h-1} d_s \, b_{h-s}.$$

The first term represents the direct promotion of the qualifying candidate, while the second term captures trickle-down effects generated by the additional female faculty present in earlier years.

An important feature of this formulation is that it automatically incorporates the catch-up dynamics observed at the individual level. When initially unsuccessful candidates are eventually promoted, the stock difference between treated and control departments shrinks, which is reflected in negative values of d_h. The resulting spillover effects are therefore net of any catch-up among initially unsuccessful candidates.

G.4 Empirical implementation

We use the estimated sequences $\{\Delta_h\}$ and $\{Y_h\}$ to construct the annual flows $\{d_h\}$ and $\{y_h\}$. Using these objects, we back out a stylized spillover kernel b_j.

Because the unrestricted sequence $\{b_j\}$ is not precisely identified at each horizon, we estimate a parsimonious parameterization in which the spillover effect follows a delayed geometric decay:

$$b_j = \begin{cases} 0 & j < L, \\ \beta\rho^{j-L} & j \geq L. \end{cases}$$

This representation captures two key features suggested by the reduced-form evidence: first, that spillovers do not emerge immediately following the promotion of the focal candidate, and second, that once they begin they persist over multiple years.

The parameters (β, ρ, L) are chosen to minimize the distance between the observed department-level cumulative effects and the cumulative effects implied by the model.

G.5 Results

The specification that best matches the reduced-form department effects implies a delay of approximately five years before spillovers begin to materialize. Conditional on this delay, the estimated spillover parameters are approximately

$$\hat{\beta} \approx 0.20, \qquad \hat{\rho} \approx 0.82.$$

These estimates imply that once spillovers begin, the promotion of one additional female faculty member generates roughly 0.20 additional female promotions in the first year of the spillover phase, with the annual effect declining gradually over time at a rate of about 18% per year. This timing is consistent with mechanisms operating through the academic pipeline, such as mentoring, influence over promotion committees, and the gradual strengthening of the female promotion pipeline within departments.

Under this parameterization, the cumulative long-run spillover generated by one additional female promotion is approximately

$$\frac{\hat{\beta}}{1-\hat{\rho}} \approx \frac{0.20}{1-0.82} \approx 1.1.$$

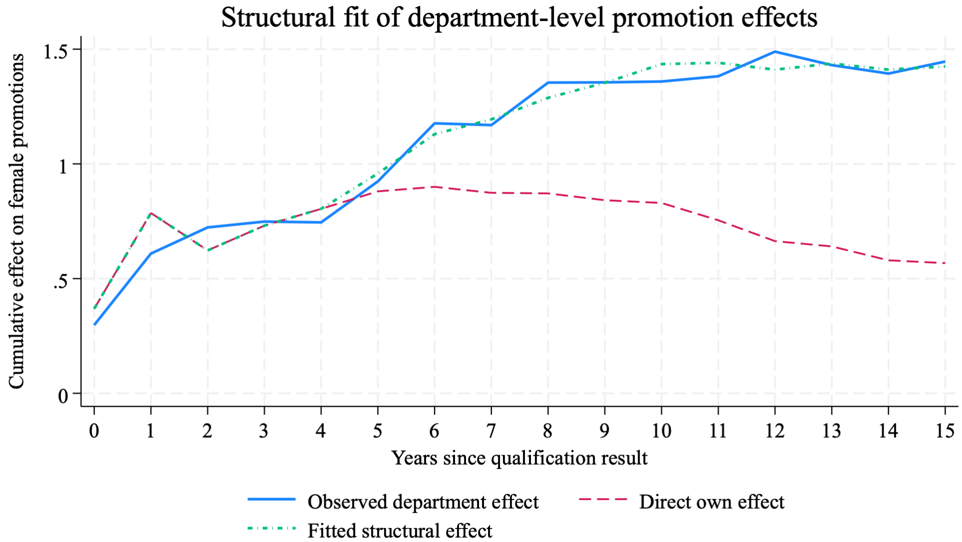
This estimate is broadly consistent with the reduced-form evidence reported in the main text. The cumulative department-level effect after fifteen years is approximately 1.45 additional female promotions, while the cumulative direct promotion effect of the qualifying candidate herself is approximately 0.57. The difference between these two quantities, roughly 0.9, reflects downstream promotions generated through departmental spillovers.

Figure G1 illustrates the fit of the structural model relative to the reduced-form department-level estimates.

G.6 Interpretation

This exercise should be interpreted as a descriptive decomposition. Its purpose is to combine the individual and department-level evidence in order to better understand the timing and magnitude of the spillover process. Nevertheless, the results reinforce the interpretation that the long-run department-level effects documented in the main text are driven primarily by propagation through departmental promotion dynamics rather than solely by the direct promotion of the qualifying candidate.

Figure G1: Propagation fit of department-level promotion effects



Notes: The figure compares the observed cumulative department-level effect of an exogenous female qualification to Associate Professor on subsequent female promotions with the fitted values from the spillover model described in Appendix G. The red dashed line shows the cumulative individual-level promotion effect of the qualifying candidate herself. The fitted path combines this direct effect with a delayed geometric spillover kernel, whose parameters are chosen to best match the observed department-level reduced form.

Appendix H — Data Sources and Data Cleaning

We collected data from several sources: (i) the Spanish Ministry of Education (TESEO) for PhD theses; (ii) the Spanish State Bulletin (BOE) for academic promotions; (iii) the Ministry of Science and Innovation for the centralized qualification exams (Zinovyeva and Bagues, 2015); and (iv) OpenAlex, Scopus, and Dialnet for research production. Below, we describe the procedures used to harmonize data across these sources.

H.1 Data sources

Spanish State Bulletin. We compile information on promotions to Associate and Full Professor positions in Spanish public universities from the Spanish State Bulletin (*Boletín Oficial del Estado*, BOE). Each public announcement reports the identity of the promoted individual, the university, and the research field, which we use to reconstruct the evolution of faculty composition.

Between 1986 and 2023, we observe approximately 68,000 promotions across 48 universities and 188 research fields. To account for promotions prior to 1986, which are only sporadically recorded in BOE, we supplement our data with the official rosters of Full and Associate Professors eligible to serve on qualification committees between 2002 and 2008. These lists include the appointment dates of faculty who have met the minimum research requirements and cover approximately 85 percent of Full Professors and 70 percent of Associate Professors.

Using these records, we construct a longitudinal panel of departments (field $\times$ university). For each department-year, we track faculty composition by academic rank and gender.

Registry of PhD Theses. Information on doctoral training comes from the official registry of PhD theses maintained by the Spanish Ministry of Education (*Teseo*).[25] The database reports the identity of PhD graduates, advisors, committee members, graduation dates, and research fields. Research fields are self-reported at submission and follow the UNESCO

25. Available at https://www.educacion.gob.es/teseo. Registration is mandatory and is estimated to cover approximately 90% of all PhD theses defended in Spain (Fuentes-Pujol and Arguimbau-Vivo, 2010).

classification system, which includes 24 two-digit fields, 250 four-digit subfields, and 2,153 six-digit specializations.[26]

National qualification exams. We use administrative data from the centralized national qualification exams conducted between 2002 and 2008 under the *habilitación* system. These exams were administered by the Spanish Ministry of Education, which made detailed administrative records publicly available. The information was archived following the exams. The records include the universe of candidates and eligible evaluators, the random draws determining the composition of evaluation committees, and the final qualification outcomes.[27]

Bibliographic databases. We measure research output using three complementary bibliographic databases: OpenAlex, Scopus, and Dialnet.

OpenAlex is a comprehensive and fully open catalogue of global scholarly production (Priem et al., 2022). It draws on multiple sources to compile comprehensive metadata on academic publications. These include journal articles, books, conference papers, and research datasets. For each item, we extract information such as title, complete author list, publication year, venue, and institutional affiliation, which we use to build individual publication records. A key strength of OpenAlex is its inclusion of content not indexed by Scopus, particularly in less formal or emerging publication outlets. Its graph-based structure linking authors, works, institutions, and concepts enables detailed mapping of co-authorships and academic careers.

Scopus is a large-scale, multidisciplinary abstract and citation database launched by Elsevier in 2004, widely used to benchmark international academic output. It offers structured metadata on articles published in indexed journals, providing information on publication years, citation counts, journal-level indicators, and author affiliations (Burnham, 2006). For our purposes, Scopus is particularly valuable because of its reliable author disambiguation, especially in cases where many researchers share similar names and publish frequently.

26. See https://skos.um.es/unesco6.

27. These data were originally collected and assembled by Zinovyeva and Bagues, 2015 and Bagues et al., 2017, who provide a detailed description of the data.

Dialnet is a bibliographic database developed by the University of La Rioja in 2001 to increase the international visibility of Spanish academic production, particularly in the humanities and social sciences (Garrido Ardila, 2020). It is now viewed as Spain's principal scientific bibliographic repository. Its coverage includes journal articles, books, book chapters, and theses, and it plays a crucial role in capturing academic activity not indexed in international databases like OpenAlex and Scopus.

H.2 Identifying researchers

To link researchers across data sources, we implement a two-step disambiguation procedure that assigns a unique identifier to each individual while minimizing the risk of conflating distinct researchers.

In the first step, we disambiguate author identities across Habilitación, BOE, and Teseo using a fuzzy-matching algorithm tailored to Spanish naming conventions. The algorithm is augmented with auxiliary information on institutional affiliation, field of specialization, year of graduation, and shared identifiers (if any) across databases. A central feature of the algorithm is its differential treatment of common Spanish forenames and surnames to account for recurring naming patterns. For example, a widespread surname like *Fernandez* in a name like 'Manuel Bagues Fernandez' is likely to be uninformative. In contrast, widespread names like *Jose* or *Pilar*, used as second names, such as 'Maria Jose' and 'Maria Pilar', are informative.

In the second step, after uniquely identifying researchers across Habilitación, BOE, and Teseo, we disambiguate publication records. We apply a parallel matching strategy that combines Spanish naming conventions with information on institutional affiliation, field of specialization, year of graduation, and individual publication trajectories.

Step 1. Disambiguating Habilitación, BOE, and Teseo

Habilitación, BOE, and Teseo often come with a unique identifier for each researcher. The goal of this step is to assign a unique identifier to researchers who appear in the data without one. It works in two main phases: first, it tries to match unidentified researchers to already-identified ones. For example, *Manuel Bagues* may appear with ID 12345 in Habilitación and Teseo, but it may lack an identifier in BOE. Second, it matches unidentified researchers among themselves. For example, *Manuel Bagues* may lack an ID in each of the data sources. A final phase resolves transitive links to ensure consistent identifiers.

Phase 0 - Cleaning of names Before any matching, researcher names are standardized through a cleaning program that converts all text to uppercase ASCII (stripping accents), removes punctuation and special characters, expands common Spanish abbreviations (e.g., "Mª" → "MARIA", "FCO" → "FRANCISCO", "FDZ" → "FERNANDEZ"), corrects known OCR-like typos where the digit 0 was read as the letter O and vice versa (e.g., "J0SE" → "JOSE", "ANTONI0" → "ANTONIO"), removes Spanish particles and connectors such as "DE LA", "DE LOS", "DEL", "Y", "I", etc., and strips titles (e.g., "DR", "DON", "DONA").

Phase 1 - Matching unidentified to identified researchers We take two input files: a list of researchers without an identifier and a list of researchers with a known identifier. We proceed as follows:

1. *Exact string matching.* The concatenated name-surname string is matched exactly between the two files. About 76% of matches are obtained at this step.
2. *Fuzzy matching - unilateral.* We first modify names in the unidentified file and then modify names in the identified file. We call a battery of sub-procedures that try to match records by systematically transforming one side's name string. The transformations include: prepending a common first name (gender-specific, drawn from a predefined list of frequent Spanish forenames) to a name that appears to lack one; removing a common first name from the full name string; dropping the second or first word from names with four or more words; constructing alternative merge keys using only the first name plus the full surname, the full name plus the first surname, or the first name plus the first surname; and adding a common second name to records that have a single given name.[28]

28. We collected the 50 most common forenames in Spain from the Spanish Statistical Office (INE). The

3. *Fuzzy matching - bilateral.* We then apply our transformations to both sides simultaneously. These include: dropping the second word from both files' name strings (with a check that the first letter of the dropped words agrees and that either one is an initial or their edit distance is at most 2); keeping the first name plus full surname on both sides; various asymmetric combinations where one side uses the full name with only the first surname and the other uses the first name with the full surname; and matching on name initials plus full surname (reducing each given name to its initial letter and joining on the resulting string, then validating that the full first names are compatible via edit distance ≤ 2, with an exemption for initials).

4. *Safeguards against false matches.* Throughout, several safeguards are in place to limit false positives. Matches involving researchers for whom both surnames belong to the list of the most frequent Spanish surnames are dropped when the match relies on partial name information, since common surname combinations are too ambiguous.[29] Gender is used as a blocking variable in the "add common first name" step (female names are only prepended to female records and vice versa). When multiple matching methods produce the same pair, exact string matching is given priority.

After all matching rounds, the candidate pairs are appended, deduplicated with priority given to exact matches, and merged back to the unidentified list. Records that were successfully matched receive the known identifier; those still unmatched are carried forward to Phase 2.

Phase 2 - Matching unidentified researchers among themselves Researchers not linked to any known identifier in Phase 1 are now matched against each other. The unidentified records that found no exact match to any other unidentified record are then matched against

list includes: {Carlos, Jesus, Alejandro, Miguel, Rafael, Angel, Pablo, Pedro, Sergio, Antonio, Manuel, Jose, Francisco, David, Juan, Javier, Daniel, Fernando, Jorge, Alvaro, Adrian, Diego, Luis, Carmen, Ana, Pilar, Laura, Josefa, Isabel, Dolores, Teresa, Marta, Cristina, Angeles, Lucia, Francisca, Antonia, Sara, Paula, Elena, Luisa, Raquel, Rosa, Manuela, Concepcion, Maria}

29. The list of common surnames, collected from the Spanish Statistical Office (INE), includes: {Fernandez, Rodriguez, Garcia, Martinez, Hernandez, Lopez, Gonzalez, Sanchez, Perez, Gomez, Martin, Jimenez, Ruiz, Diaz, Moreno, Munoz, Alvarez, Romero, Gutierrez, Navarro, Torres, Dominguez, Ramos, Vazquez, Ramirez, Gil, Serrano, Morales, Blanco, Suarez, Castro, Ortega, Delgado, Ortiz, Marin, Rubio, Nunez, Medina, Iglesias, Castillo, Cortes, Garrido, Santos, Guerrero, Lozano, Cano, Alonso, Molina, Sanz, Mendez, San}.

the full pool of remaining unidentified researchers using the same unilateral and bilateral fuzzy-matching routines from Phase 1. The unilateral routines are applied in both directions. Self-matches are dropped.

Phase 3 - Resolving Transitive Links, Validating Fuzzy Matches, and Assigning Final Identifiers

1. *Resolving transitive links.* The pairwise matches produced in Phase 2 can generate transitive chains (e.g., A is matched to B and B is matched to C). These are consolidated into groups by iterating over unidentified identifiers, collecting all directly linked partners, assigning the group a single representative identifier, and removing processed records before continuing. Records that were never matched to any other unidentified researcher are retained as singletons. The output is a file in which each cluster of transitively linked unidentified records shares a common identifier.
2. *Validating fuzzy matches.* The grouped matches among unidentified researchers are evaluated analogously but more simply. Records that appear only once (i.e., were never linked to another unidentified researcher) are accepted as singletons. For the rest, the merge key is again decomposed into common and uncommon components. Matches are accepted if the uncommon portion is substantive. Among the remaining cases, those where the linked records share the same discipline are accepted. The few residual cases are manually inspected and retained.
3. *Assigning final identifiers.* The three outputs—researchers with known identifiers (updated after the validation step above), previously unidentified researchers now matched to known identifiers, and previously unidentified researchers matched only among themselves—are combined into a single table. The three outputs from the previous steps are combined and mapped to a single sequence of integer identifiers. This identifier is then merged back into each source dataset – PhD theses, promotions, and national qualification records – by name string and event identifier, which are unique. An extensive set of manual corrections resolves cases that the automated procedure could not handle, predominantly researchers whose names are composed entirely of high-frequency Spanish forenames and surnames.

These fixes typically disambiguate by conditioning on discipline, university, or national identity number. Consistency checks flag likely false merges – mainly identifiers mapping to more than two distinct birth years – and split them into separate individuals. Thesis supervisors and committee members, who were not part of the main disambiguation, are then linked to the harmonized identifier set by matching on name, event, university, and discipline, with residual ambiguities again resolved manually. The final output is a master researcher file containing, for each researcher-event, the unified identifier along with year, age, gender, university, academic rank, discipline, and source dataset. We validated this procedure extensively through manual checks. Our margin of error is less than 0.5%.

Step 3. Merging Habilitación, BOE, and Teseo with Publication Data

We now merge our full list of researchers, observed in Habilitación, BOE, and Teseo, with the publication data. We do it separately for each data source (Dialnet, OpenAlex, Scopus).

Merging with Dialnet We link our disambiguated researchers to author profiles in Dialnet. Candidate matches are generated using the same battery of exact and fuzzy name-matching routines applied in earlier stages. Each candidate pair is then augmented with auxiliary information from both sides: from Dialnet, the author's university affiliation, department-based discipline, academic area, and window of active publication years; from our administrative records, universities of promotion, PhD university, habilitacion university, area, and discipline. Disambiguation proceeds through a number of sequential steps of decreasing strictness. The earliest steps require agreement on university affiliation, academic area or discipline, an exact or uncommon-name match, and a plausible window of Dialnet activity relative to the researcher's PhD or career dates. Subsequent steps progressively relax these conditions, first dropping the affiliation requirement when the Dialnet profile lists no university, then accepting broader discipline categories (four-category, then STEM/non-STEM), then accepting matches over common names provided the match is one-to-one and no available variable disagrees. For one-to-one matches, the procedure eventually accepts cases based solely on a plausible activity window combined with no conflict in affiliation or discipline information. For one-to-many matches—where a single researcher maps to multiple Dialnet profiles or vice versa—the procedure requires stronger conditions throughout, typically demanding agreement on affiliation or discipline plus uncommon surnames and more than one published paper. At each step, once a researcher–profile pair is confirmed, competing candidates for that researcher are dropped before proceeding. After all steps, remaining many-to-one or one-to-many links are resolved by retaining the candidate with matching discipline or area, and a handful of manual corrections address residual conflicts.

Merging with Scopus We link our disambiguated researchers to author profiles in Scopus by first constructing candidate links through targeted Scopus lookups that expand each researcher's name into plausible variants and then merging Scopus outputs back to the administrative backbone (Habilitación candidates and evaluators, BOE promotions, and Teseo PhD records). For each candidate researcher-Scopus pair we enrich both sides with auxiliary information: on our side, university affiliation, discipline and area, promotion/qualification years, PhD graduation year, and year of birth; on the Scopus side, author name strings, paper-level publication years, affiliation strings taken from the first reported affiliations on each paper, and paper disciplines inferred from MAG-based codes. Disambiguation proceeds as a sequence of increasingly permissive checks, always conditioning on a plausible publication timeline (papers can precede the PhD by up to five years when a thesis is observed; otherwise papers must begin after age 20 when year of birth is known, and in all cases the implied career span is capped). Matches are accepted when corroborating evidence accumulates across a researcher's papers, prioritizing (i) overlap between administrative universities and Scopus affiliations combined with exact full-name matches and consistent STEM/non-STEM classification, (ii) affiliation overlap combined with agreement in four-category discipline at least once across papers, and (iii) for thesis-only and candidate records (where mobility is more likely), exact-name matches coupled with discipline agreement and "uncommon-name" indicators derived from filtering out frequent surnames/forenames. After a match is confirmed, competing candidates are dropped to enforce a one-to-one mapping between our researcher ID and Scopus identifier. Residual many-to-one or one-to-many cases are resolved in post-processing by retaining the link that best fits event timing (publication years relative to promotion/qualification), maximizes overlap between area labels and MAG/journal text, prefers richer name information, and (as a final tie-breaker) has the greatest number of matched papers; a small number of remaining conflicts are handled manually, yielding a final crosswalk between our researcher identifiers and Scopus author IDs used to assemble paper- and author-paper level Scopus panels.

Merging with OpenAlex We link our disambiguated researchers to author profiles in OpenAlex. The procedure applies the same battery of exact and fuzzy name-matching routines used in the researcher disambiguation to generate candidate matches between our researcher names and OpenAlex author names. These candidate pairs are then pre-processed: matches where available ORCID or Scopus identifiers conflict across the two sides are dropped, and name conflicts are identified by checking each word of the OpenAlex author name against all known versions of the researcher's name in our data (with edit distance tolerances for typos). For the remaining candidates, we augment both sides with auxiliary variables useful for disambiguation. On the OpenAlex side, we extract each author's Spanish public university affiliations, publication activity years, and the ISSNs of their top journals. On our side, we construct analogous variables from the administrative records (university of promotion, university of PhD graduation, affiliation at the time of the qualification exam, and discipline) and from the Spanish bibliographic database Dialnet (journal ISSNs and institutional affiliations). Disambiguation then proceeds in three sequential tiers ordered by the informativeness of the name match: exact string matches first, then substring-based fuzzy matches (low information loss), and finally initial-based matches (high information loss). Within each tier, candidate pairs are confirmed using the same hierarchical set of checks: first, agreement on ORCID or Scopus identifiers; then, agreement on institutional affiliation combined with discipline; then, overlap in journal ISSNs between Dialnet and OpenAlex; then, narrow discipline agreement; and finally, plausibility of the OpenAlex activity period relative to PhD completion. Matches confirmed in earlier tiers are removed before processing later tiers to ensure each OpenAlex profile is assigned to at most one researcher. Researchers matched to more than ten OpenAlex profiles who have common East Asian name patterns or no Spanish university affiliation in OpenAlex are dropped. At the compilation stage, any OpenAlex profile assigned to multiple researchers across tiers is resolved in favor of the higher-priority tier. A final validation step drops matches for thesis-only researchers whose OpenAlex activity window is implausible relative to the PhD date and who lack any corroborating disambiguation variable. Where a single researcher is linked to multiple OpenAlex profiles, these are grouped into a single composite identifier.

Appendix Table H1 shows the share of researchers matched across different databases.

Table H1: Identified researchers across databases

	PhD	Candidates	Promotions	Supervisors	PhD Committees	Exam Evaluators	Publications
PhD	100%	5%	15%	21%	32%	7%	77%
Candidates	77%	100%	73%	76%	87%	37%	94%
Promotions	74%	23%	100%	68%	82%	39%	93%
Supervisors	80%	19%	54%	100%	88%	36%	92%
PhD Committees	83%	15%	45%	61%	100%	28%	89%
Exam Evaluators	58%	21%	71%	81%	92%	100%	91%

Notes: The table reports the share of researchers identified across different databases. Information on PhD graduates, supervisors, and committee members comes from TESEO. Data on candidates and exam evaluators comes from the Ministry of Science and Innovation. *Promotions* refers to the complete list of individual promotions to Associate and Full Professorships between 1986 and 2023, collected from the Spanish State Bulletin. Each row shows the share of individuals in a specific group who appear in each database. For example, 77% of candidates appear in TESEO as PhD graduates, 76% as PhD supervisors, 87% as PhD committee members, and 94% appear in the research output database.

H.3 Imputing affiliations in Habilitación

We address a key limitation of the *Habilitación* records: candidates' institutional affiliations are not directly observed. We therefore reconstruct affiliations using a sequential imputation strategy that prioritizes administrative sources and then complements them with publication-based evidence. First, we assign affiliations using the identifiers-based database that links candidates to prior academic events. In particular, full-professor candidates also appear as evaluators in associate-professor exams, allowing us to recover their university at the time they served as evaluators. For early-career candidates who cannot serve as evaluators and do not yet appear in the BOE promotions records, we use the PhD institution when the doctorate is awarded in the exam year or the preceding year. Second, we exploit candidates' research output in Scopus/OpenAlex in the five years prior to the exam (and, when available, the year immediately following). We interpret these affiliations as reflecting candidates' institutional attachment at the time of the *Habilitación* exam and assign the Spanish public university affiliation observed for their publications closest to the exam date. When publications list more than one affiliation within this window, we retain the affiliation associated with the largest number of papers (and, in the rare cases with multiple Spanish public university affiliations, the one with the strongest presence in the years closest to the exam). When a publication reports both a public university and a research centre, we keep the university affiliation, restricting institutions to our harmonized list of 48 Spanish public universities. Finally, if neither administrative sources nor publications yield a match, we use Dialnet as a last-resort "top-up" source, which reports researchers' most recent institutional affiliation; we only retain this information when it corresponds to a Spanish public university, and otherwise code the affiliation as private/abroad.

H.4 Title deduplication

For our purposes, we see OpenAlex, Scopus, and Dialnet as three complementary data sources. OpenAlex is a comprehensive and fully open catalogue of global scholarly production, which we use as the foundation of our publication records. Scopus is particularly valuable for its reliable author disambiguation. Dialnet is Spain's principal scientific bibliographic repository, crucial to capture academic activity not indexed in OpenAlex and Scopus.

We deduplicated publication records across the three databases to avoid double-counting individuals' research output. To do so, we follow a multi-step procedure that identifies and clusters duplicate titles across sources based on string similarity. First, we normalize all titles by lowercasing and removing punctuation, diacritics, LaTeX formatting, and HTML tags. We also convert Greek characters to their textual equivalents and remove stopwords in both English and Spanish. This step improves the accuracy of subsequent fuzzy matching by minimizing spurious textual variation.

After normalization, we apply a multi-step fuzzy matching algorithm to detect likely duplicates. We compute similarity scores between titles using *token_set_ratio* and *partial_ratio* from the Python *RapidFuzz* library. We perform multiple rounds of fuzzy matching across publication records, grouping by individual ID, publication year (allowing a $\pm$1-year window to accommodate discrepancies across databases), and publication identifier. The matching proceeds through increasingly flexible grouping criteria: early steps group by exact ISSN; intermediate steps cross-match print and electronic ISSNs between OpenAlex and Scopus to account for inconsistent ISSN assignments across databases; and later steps relax the ISSN requirement entirely for records where one source reports a missing identifier. We apply similarity thresholds of 98, 100, and 90 across these steps. Each step identifies clusters of duplicate titles. Within each cluster, we retain a single representative record, prioritizing sources in the following order: OpenAlex, Scopus, then Dialnet.

Because the pairwise deduplication can produce transitive chains (e.g., record A maps to B, which maps to C), we run a post-processing step that identifies connected components in the resulting mapping graph and selects a single canonical identifier per component based on the highest citation count. All mappings from original to deduplicated *work_id* values are stored and used to compute publication-based productivity measures. Unmatched records retain their original identifier.

Appendix Table H2 summarizes the final publication dataset. After deduplication, we observe over six million unique publication records: 72% are journal articles, 17% books or book chapters, and the remainder conference proceedings or other formats. Because we assign priority to OpenAlex when deduplicating records across sources, it accounts for 50% of the total coverage. Scopus and Dialnet contribute an additional 25% and 25% of unique entries, respectively. These additions are unevenly distributed across fields. Scopus primarily adds content in physical and health sciences, while Dialnet contributes disproportionately to humanities and social sciences, and especially so in Spanish-language publications.

In total, we link publications to 214,431 unique individuals. Appendix Table H1 shows that, among all individuals in our data, 77% of PhD graduates, 93% of faculty members, 94% of exam candidates, and 91% of evaluators are matched to at least one academic output.

Table H2: Publication database

	Total	Health Sciences	Life Sciences	Physical Sciences	Social Sciences	% Spanish
Number of publications	6,430,173	26.56%	18.65%	33.52%	21.28%	34.98%
Number of publications by type						
Journal articles	4,567,444	26.89%	18.88%	32.29%	21.94%	30.48%
Books/Book chapters	1,098,931	13.78%	13.57%	39.67%	32.98%	74.12%
Conference proceedings/other	763,798	29.58%	19.02%	41.36%	10.03%	5.60%
Number of publications by source						
OpenAlex	3,220,647	28.03%	20.10%	30.35%	21.53%	23.81%
Scopus	1,607,887	23.10%	15.45%	50.75%	10.71%	3.60%
Dialnet	1,601,639	16.85%	8.32%	16.15%	58.68%	88.96%

Notes: The table reports descriptive statistics for publication records, after deduplication across OpenAlex, Scopus, and Dialnet. Column 1 shows the total number of unique entries. Columns 2–5 present the distribution across four broad research domains, following OpenAlex's research topic taxonomy (see Appendix H.5). Column 6 reports the share of publications in Spanish.

H.5 Research topic classification

In OpenAlex, each publication record comes with both a set of *topics*, from the OpenAlex topic taxonomy, and a set of *concepts* inherited from the Microsoft Academic Graph (Wang et al., 2019; Hafner and Hedtrich, 2024). Scopus, Dialnet, and Teseo records lack this information, so we run the Microsoft Academic Graph (MAG) text classifier on English titles and abstracts.[30] We infer topics from MAG concepts by estimating $P(\text{topic} \mid \text{concept})$ on OpenAlex data and then projecting this mapping to Scopus, Dialnet, and Teseo.

The OpenAlex taxonomy provides a hierarchical classification of scholarly topics. It consists of four domains (e.g., *Social Sciences*), subdivided into 26 fields, 252 subfields, and 4,516 topics (e.g., *Labor market dynamics and wage inequality*). Appendix Table H3 shows the mapping from domains to fields. The complete taxonomy is available at https://api.openalex.org/topics.

We start from OpenAlex work–topic–concept triples. We keep concept assignments with a

30. We use English-translated titles for Dialnet.

Table H3: OpenAlex Taxonomy

Health Sciences	Life Sciences
- Dentistry	- Agricultural and Biological Sciences
- Health Professions	- Biochemistry, Genetics and Molecular Biology
- Medicine	- Immunology and Microbiology
- Nursing	- Neuroscience
- Veterinary	- Pharmacology, Toxicology and Pharmaceutics
Physical Sciences	**Social Sciences**
- Chemical Engineering	- Arts and Humanities
- Chemistry	- Business, Management and Accounting
- Computer Science	- Decision Sciences
- Earth and Planetary Sciences	- Economics, Econometrics and Finance
- Energy	- Psychology
- Engineering	- Social Sciences
- Environmental Science	
- Materials Science	
- Mathematics	
- Physics and Astronomy	

Notes: The table lists 26 research fields, each mapped into one of four broad research domains defined by the OpenAlex taxonomy.
Source: https://api.openalex.org/topics.

concept score ≥ 0.4 and aggregate OpenAlex topic scores within concept–topic pairs.

$$\text{score}_{c,t} = \sum_{w \in \mathcal{W}_{c,t}} \text{topic score}_{w,t}$$

where $\mathcal{W}_{c,t}$ is the set of works w that contain concept c and topic t. For each concept c, we normalise to obtain

$$P(t \mid c) = \frac{\text{score}_{c,t}}{\sum_{t'} \text{score}_{c,t'}}$$

To reduce noise, we retain the top 10 highest-$P(t \mid c)$ topics per concept. Let a given work i have MAG concept set $\mathcal{C}_i$ with classifier scores $s_{i,c} \in [0,1]$. We merge each concept $c \in \mathcal{C}_i$ to its concept–topic probabilities $P(t \mid c)$ and form a work-specific score for topic t as follows.

$$\widetilde{p}_{i,t} \; = \; \sum_{c \in \mathcal{C}_i} s_{i,c} \cdot P(t \mid c).$$

We then normalise across topics for that work,

$$p_{i,t} \;=\; \frac{\widetilde{p}_{i,t}}{\sum_{t'} \widetilde{p}_{i,t'}},$$

and assign one topic per work as $\widehat{t}_i \;=\; \arg\max_t \, p_{i,t}$.

For example, the MAG concept *Game Theory* maps with high $P(t \mid c)$ to OpenAlex topics like *Evolutionary Game Theory and Cooperation*, *Game Theory and Applications*, and *Auction Theory and Applications*. For a Scopus (or Dialnet/Teseo) record with a strong *Game Theory* concept score, the work-level distribution $p_{i,t}$ is constructed by weighting these candidate topics by both $P(t \mid \text{concept})$ and the work's concept scores. The assigned topic is the single topic with the highest $p_{i,t}$.