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

Manuel Bagues Milan Makany Giulia Vattuone Natalia Zinovyeva*

August 19, 2026

Abstract

We study how promotion decisions shape female scientists' careers and the academic pipeline, using data from 4,000 Spanish university departments. In Spain's centralized *habilitación* system, national qualification committees effectively controlled access to senior academic ranks during 2002–2008. We identify exogenous variation in promotions using the random assignment of evaluators to these committees: applicants evaluated by a coauthor or colleague were more likely to qualify and be promoted. We document two main findings. First, failing at this career gate has more persistent consequences for women: female Associate Professor candidates who narrowly miss qualification are 57 percentage points less likely to be promoted fifteen years later, compared with 29 for men. Second, when women qualify, the effects extend beyond their own careers: their departments subsequently promote more women, graduate more female PhD students, and produce more female PhD graduates who remain active in research and advance in academia.

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 Catherine Beaudry, 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, Tinbergen Institute, WOEPS, University of Zurich, KTI-KRTK, University of Bath, University of Vienna, Diversity and Human Capital Workshop at the University of Exeter, and the RFBerlin-CEPR Annual Symposium. Natalia Zinovyeva gratefully acknowledges support from the CAGE Research Centre, funded by the ESRC.

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 PhD 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 persist even in fields with relatively higher female representation: in the Social Sciences and Humanities, women hold less than 40% of full professorships. Economics presents one of the starkest gender gaps: women account for only 23% of full professors in Europe and 18% in the US (Auriol et al., 2020; Tesar, 2025). Despite decades of growth in women's doctoral attainment, these gaps remain remarkably persistent, consistent with the so-called "leaky pipeline", a metaphor for women's disproportionately high attrition and slower advancement across 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).

Gender gaps at senior ranks are often attributed, at least in part, to barriers that accumulate over academic careers. Women may face unequal access to mentoring, networks, funding, publishing opportunities, visibility, or credit in evaluations. Such barriers can affect who reaches the promotion margin and how researchers are evaluated once they do. They also provide a rationale for policies aimed at increasing women's representation in senior academic positions, as a way to offset disadvantages accumulated early in female researchers' careers.[1]

This paper focuses on two dynamic mechanisms that can amplify or mitigate these barriers

1. Examples include women-only professorships, gender-balance targets in hiring and promotion, and funding schemes tied to the hiring of female researchers. Women-only full-professor posts have been opened in several countries including Austria, Australia, Ireland and the Netherlands, for instance at the University of Adelaide (https://www.abc.net.au/news/2018-08-22), Eindhoven University of Technology (https://www.science.org/content/article/men-need-not-apply-university-set-open-jobs-just-women), the Vienna University of Technology (https://informatics.tuwien.ac.at/news/1429), and the University College Dublin (https://www.siliconrepublic.com/innovation/women-only-professor-positions-sali-ireland). In Germany, certain grants have been tied to the hiring of women since 2012 (Wissenschaftskonferenz, 2016).

over time. The first concerns how researchers respond to setbacks at key career stages. Academic careers require researchers to repeatedly decide whether to keep investing while future prospects remain uncertain. A failed promotion may therefore do more than delay advancement: it may change beliefs about future prospects, discourage reapplication, or push researchers toward outside options. If these responses differ by gender, because women and men face different constraints, outside options, expectations, or behavioral responses to failure, promotion systems can gradually reduce the pool of women available for senior positions even when women enter the pipeline in large numbers. The second mechanism is spillovers from senior representation. Women in senior positions may make academic careers more attractive and sustainable for junior women through mentoring, role-model exposure, networks, workplace climate, or access to more appealing research agendas. Senior female representation may therefore be both an outcome of past barriers and an input into the future evolution of the academic pipeline.

We provide causal evidence on these two mechanisms. First, we ask whether promotion outcomes have lasting effects on researchers' attachment to academia and research activity, and whether these effects differ by gender. Second, we ask whether promoting a woman to a tenured faculty position changes the long-run gender composition of her department and affects the academic outcomes of future cohorts of women, as reflected in female representation among PhD graduates and their subsequent academic trajectories.

To answer these questions, we exploit a natural experiment generated by the Spanish national academic qualification system, *habilitación*, in place between 2002 and 2008. During this period, candidates for Associate and Full Professor positions in public universities could be promoted by their university only after obtaining a national qualification from a field-level committee. The qualification stage was highly selective: the number of available qualifications was capped, and roughly 10% of candidates qualified. This cap sharply reduced the number of promotions relative to both the preceding decentralized promotion system and the subsequent system, in which the qualification stage was uncapped and no longer binding. The *habilitación* therefore created a binding bottleneck for advancement. Qualification decisions were highly consequential: more than 90% of successful candidates were promoted within three years.

The main source of variation comes from the random composition of evaluation committees. Candidates first applied to a field-level qualification exam. Once applications closed, committees were selected by lottery from the pool of eligible evaluators in that field.[2] In the absence of conflict-of-interest rules excluding evaluators with prior academic ties to candidates, some candidates were randomly assigned committees that included a strong connection: a coauthor, colleague, or PhD advisor. As Zinovyeva and Bagues (2015) show, around one third of candidates were assigned such a connection, and having one increased qualification chances by about 40 percent. Our empirical strategy compares candidates who, given the pool of eligible evaluators and the rules of the draw, had the same expected number of connections but differed in whether the lottery actually assigned them a connected evaluator. The lottery therefore shifted promotion outcomes for candidates at the margin of qualification.

The setting is uniquely suited to studying the dynamic consequences of promotion. First, committee assignment generates quasi-random variation in promotion outcomes among candidates in the same field and exam. Second, the *habilitación* created a restrictive and publicly scrutinized bottleneck. Because qualifications were tightly capped, candidates whose success depended on a connected evaluator were not weak candidates promoted in the absence of merit, but researchers near the promotion margin in a highly selective national process. Consistent with this interpretation, we show that these marginally promoted candidates had strong academic records, substantially closer to candidates who qualified regardless of committee composition than to candidates who did not qualify even with favorable committee draws. Third, the system was short-lived. After 2008, it was replaced by a less restrictive procedure with an uncapped and non-binding qualification stage, allowing us to follow candidates and departments after the initial shock in an environment closer to the standard university promotion regime. Yet the financial crisis and subsequent promotion freeze limited immediate faculty adjustment, so effects on subsequent faculty promotions are likely to unfold gradually.

2. There were one to three evaluation rounds per year. In each round, a new committee was drawn for each field-rank cell and evaluated all candidates applying in that cell. For example, separate committees were drawn for Associate Professor and Full Professor candidates in several economics fields, including Applied Economics, Economic History, and Economic Analysis.

We combine this source of variation with novel administrative data covering around 4,000 Spanish university departments across all disciplines. The data link the full universe of qualification exams – including candidates, potential evaluators, and actual committee members – to a long historical panel of promotions to senior academic ranks, the universe of PhD dissertations, and publication records from multiple bibliographic sources. This allows us to follow individual candidates, faculty composition, and PhD student outcomes for up to fifteen years after the initial qualification decisions.

We first examine the individual consequences of qualification. Among candidates for Associate Professor, those induced to qualify by a favorable committee draw are substantially more likely to obtain tenure within three years and remain more likely to hold a tenured position even fifteen years later. Marginal candidates who qualify because of a favorable draw are 54 percentage points more likely to hold a tenured position after ten years, and 40 percentage points more likely after fifteen years, than comparable candidates with an unfavorable draw.

Consistent with the idea that academic careers may be especially unforgiving for women when setbacks occur, these effects are substantially more persistent for female candidates. Ten years after the qualification exam, the tenure gap between marginally successful and marginally unsuccessful candidates is 83 percentage points for women, compared with 38 percentage points for men. This gender difference appears to reflect weaker persistence after an unsuccessful attempt: marginally unsuccessful female candidates are 28 percentage points less likely to reapply in future evaluation waves than marginally unsuccessful male candidates, and they also reduce research activity more sharply. These results indicate that the same negative promotion shock can have more durable consequences for women, reducing the future pool of women available for senior positions.

We then examine whether an exogenous female promotion changes the gender composition of the candidate's department. Identification comes from lottery-generated variation across departments: in some departments, women qualify at the margin after drawing more connected evaluators than expected, while in others they fail to qualify after an unfavorable draw. In the short run, an exogenous female qualification mechanically increases the number of women in senior ranks. The effect then builds over time, reaching approximately 1.5 additional female

Associate Professor promotions after twelve years, an estimate that remains statistically significant. Estimates remain positive at longer horizons, although less precisely estimated. A marginal promotion shock therefore does not merely change the career of one candidate; it generates a persistent shift in the gender composition of her department's senior faculty. This pattern is consistent with the individual-level evidence that unsuccessful women are less likely to remain in academic careers. The point estimates further suggest that female promotions may generate additional spillovers on subsequent female faculty promotions, although these estimates are noisy.

The downstream effects on PhD students are large. Each exogenously promoted woman leads to approximately six additional female PhD graduates over a ten-year horizon, four additional female graduates who are research-active five years after graduation, and one additional female graduate who reaches a tenured position within ten years of graduation. We find no comparable effects on male PhD graduates or their subsequent outcomes, nor analogous downstream effects from exogenously promoted men. The effects therefore appear to be gender-specific rather than reflecting a general expansion in doctoral training or departmental activity.

Several patterns help interpret these effects. First, impacts on female graduates' academic outcomes appear for cohorts already present in the department when the woman obtained tenure, indicating that the effects are not driven only by attracting new female students. Second, additional female graduates are not disproportionately advised by women: the share advised by female faculty is close to the average rate. This suggests that the effects are not solely due to direct female adviser–student matches, but also reflect broader department-level channels. Third, the gains in research activity and future tenure imply that the additional female graduates are not simply marginal students being pushed through the PhD with weaker academic prospects. Rather, female promotions appear to increase the number of women who remain on academic trajectories.

We cannot fully distinguish whether these effects operate by retaining high-potential female students who would otherwise have left academia, by improving the productivity and career prospects of female students already in the department, or by a combination of both. The

evidence is nevertheless consistent with a department-level mechanism in which senior women make academic careers more attractive or sustainable for junior women. Finally, the effects on both faculty composition and PhD outcomes are concentrated in departments where women are present but not yet in the majority, suggesting that spillovers may require some pre-existing female presence rather than being triggered by the promotion of a first woman in an otherwise all-male department.

We next examine whether the spillovers from female promotions operate through gender itself or through the research areas in which women are more represented. Women and men tend to work in different fields and topics, and research areas with higher female representation may attract more female students and faculty. To separate these channels, we estimate the effects of promoting researchers in more *feminized* research areas, regardless of candidate gender. These promotions also increase the number of female faculty, suggesting that research orientation can affect departmental gender composition. One interpretation is that expanding areas where women are more represented creates opportunities for future female hires or strengthens support for female advancement within the department. However, the effects on female PhD graduates are weaker and less precise, especially in the short run, than the effects of promoting a woman directly. This pattern suggests that research alignment alone is not sufficient to explain the impacts on female students. The downstream effects on PhD students appear to operate at least partly through the gender composition of senior faculty and related mechanisms, such as mentoring, role-model exposure, professional networks, and workplace climate.

This paper contributes to six strands of research. First, it contributes to the literature on barriers women face in academic careers. Prior work documents gender differences in several inputs to promotion, including the effects of formally neutral tenure policies, credit attribution, publication standards, and professional recognition (Antecol et al., 2018; Sarsons et al., 2021; Card et al., 2020; Koffi, 2021; Koffi, 2025).[3] These studies imply that women

3. Antecol et al. (2018) show that gender-neutral tenure-clock extensions in top economics departments increased male tenure rates while reducing female tenure rates. Sarsons et al. (2021) shows that female economists receive less tenure credit for coauthored work, especially when coauthoring with men. Card et al. (2020) find no evidence that referee gender affects the relative evaluation of female-authored economics papers, but show patterns consistent with a higher publication bar for women. Koffi (2021) documents gendered

may face different constraints, receive different credit, or accumulate different signals of productivity by the time they reach promotion. We show that gender inequality can also be reinforced after the promotion decision: the same negative promotion outcome leads to lower reapplication, sharper declines in research activity, and weaker attachment to academia for women than for men.

Second, this finding connects to a broader literature on gender differences in responses to competition, negative signals, and early setbacks (Buser and Yuan, 2019; Wasserman, 2023; Kang et al., 2024).[4] We add evidence from academic labor markets, where a failed promotion affects not only current rank but also subsequent research activity, reapplication, and attachment to academia. The results identify failed promotion at the tenure stage as a mechanism through which academic careers may become especially unforgiving for women, reducing the future pool of women available for senior positions.

Third, we provide large-scale causal evidence on whether senior female representation generates downstream effects on junior women. Existing evidence points in this direction but is primarily descriptive: Boustan and Langan (2019) shows that economics departments with more female faculty also have stronger female representation among students and junior faculty, although such correlations may partly reflect departmental culture, resources, or selection. We contribute to this evidence by exploiting quasi-random variation from Spain's national qualification system. Our results indicate that senior representation can shape the academic pipeline well beyond the promoted faculty member's own career.

Fourth, we contribute to understanding the mechanisms through which senior representation matters. Prior work on role models and mentoring shows that exposure to senior women can shape women's educational and academic trajectories (Carrell et al., 2010; Porter and Serra, 2020; Bettinger and Long, 2005; Ginther et al., 2020).[5] Descriptive studies also link

---

citation patterns in top economics journals, and Koffi (2025) shows that all-female papers and early-career women receive fewer citations in several fields.

4. Buser and Yuan (2019) show, using laboratory experiments and the Dutch Math Olympiad, that women are more likely than men to stop competing after losing. Wasserman (2023) documents a related pattern in electoral politics, where early losses have more persistent discouragement effects for women than for men. Kang et al. (2024) provides further evidence on gender differences in responses to negative career signals.

5. Carrell et al. (2010) show that female students at the U.S. Air Force Academy are more likely to complete

female advisers or mentors to stronger academic orientation and later outcomes among female PhD students (Hilmer and Hilmer, 2007; Gaule and Piacentini, 2018; Hofstra et al., 2022; Rossello et al., 2024; Neumark and Gardecki, 1998).[6] Other work points to broader organizational channels, including climate, harassment, and family-friendly policies (Folke and Rickne, 2022; Adams-Prassl et al., 2024; Epifanio and Troeger, 2020).[7] A separate literature asks whether female evaluators improve women's chances in hiring, promotion, or publication decisions, with mixed or limited evidence (Bagues et al., 2017; Deschamps, 2024; Card et al., 2020).[8] Our findings point to a broader department-level channel. Senior female representation matters at the department level, but the effects may not operate primarily through direct adviser–student matching or formal evaluation support. Instead, the evidence points to mechanisms such as informal mentoring, role-model exposure, professional networks, workplace climate, and research agendas.

Fifth, we distinguish the role of gender from the role of research orientation and draw implications for policies aimed at increasing women's representation in academia. Research fields and topics are not gender-neutral: women and men cluster in different areas, and male-dominated senior faculties may disadvantage candidates in areas where women are more represented if evaluators favor research closer to their own agendas (Dolado et al., 2012; Huang et al., 2020; Kozlowski et al., 2022). Recent work also shows that researcher gender composition can shape the production and direction of scientific knowledge (Bello et al., 2025; Belot et al., 2023).[9] We compare promotions of women with promotions of researchers

STEM majors when taught by female instructors; Porter and Serra (2020) find that exposure to successful women increases female economics majoring; Bettinger and Long (2005) finds mixed effects of instructor gender across fields; and Ginther et al. (2020) shows in a randomized mentoring intervention that junior female economists mentored by senior women gain in research productivity and job satisfaction.

6. Hilmer and Hilmer (2007), Gaule and Piacentini (2018), Hofstra et al. (2022), and Rossello et al. (2024) generally find stronger subsequent outcomes among female PhD students with female advisers or mentors, although Neumark and Gardecki (1998) finds mixed evidence.

7. Folke and Rickne (2022) and Adams-Prassl et al. (2024) show that women in managerial leadership are associated with less sexual harassment and better organizational climate, while Epifanio and Troeger (2020) argues that female representation can shape family-friendly policies.

8. Bagues et al. (2017) find that female applicants in Spain and Italy do not benefit from being evaluated by more women, Deschamps (2024) reaches a similar conclusion in French academic hiring committees, and Card et al. (2020) finds no evidence that referee gender affects the evaluation of female-authored economics papers.

9. Bello et al. (2025) links gender representation to research similarity and academic trajectories, while Belot et al. (2023) shows that gender diversity can affect the diversity of ideas.

working in more female-represented areas. Promotions in female-represented areas increase female faculty representation, suggesting that research orientation can shift departmental gender composition. This is policy-relevant because research-priority interventions may be less vulnerable to backlash associated with explicitly gender-based affirmative action.[10] At the same time, these promotions have weaker effects on female PhD outcomes than promotions of women directly. Research orientation is therefore not a substitute for senior female representation, but may provide a complementary route to improving faculty gender balance, with downstream student effects likely depending on the senior women such policies help bring into the department.

Sixth, our findings speak to the broader literature on gender and organizations. Existing work studies how female representation in leadership affects hiring, promotion, retention, workplace culture, networks, and organizational outcomes (Kunze and Miller, 2017; Azmat and Boring, 2020; Folke and Rickne, 2022; Cullen and Perez-Truglia, 2023).[11] Academic departments provide a setting in which senior representation can affect not only contemporaneous colleagues, but also the formation of future cohorts. By following departments for up to fifteen years, we show how representation effects propagate from faculty composition to doctoral training and then to graduates' subsequent academic careers. This long-run, intergenerational perspective is difficult to observe in most organizational settings and is central to understanding how gender gaps reproduce over time.

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 outlines our identification strategy. Section 5 presents the main results, and finally Section 6 concludes.

10. Janys (2024) shows that departments may respond to diversity objectives through limited or implicit targets rather than broad transformation.

11. Kunze and Miller (2017) show that female managers are associated with better junior women's career outcomes in firms. Azmat and Boring (2020) review evidence on how gender diversity shapes organizational behavior and performance. Folke and Rickne (2022) provides evidence that female leadership can improve workplace climate, and Cullen and Perez-Truglia (2023) shows that gendered access to informal professional networks can shape career progression.

2 Institutional Context

This section describes the institutional features that are important to our empirical design. We first describe the structure of academia in Spain: the academic ranks in the university system, and the role of departments as the relevant organizational units. We then describe the *habilitación* system, which generates the quasi-random variation in promotion outcomes used in the analysis. Finally, because our downstream outcomes include PhD graduates and their subsequent academic careers, we describe how doctoral training is organized.

2.1 Academic Ranks, Fields, and Departments

Our analysis focuses on Spain's public university system, which accounted for more than 90% of faculty and students during our period of study. Academic careers in this system are organized around nationally defined ranks. The two senior civil-service ranks are Associate Professor (*Profesor Titular de Universidad*) and Full Professor (*Catedrático de Universidad*). Below these ranks are non-civil-servant positions, including tenured and non-tenured Assistant Professor positions.[12] Our analysis focuses primarily on promotion to *Profesor Titular de Universidad*, the main entry point into Spain's senior civil-service academic ranks. This transition is the closest analogue to tenure decisions in the United States: it represents a major step in academic careers and provides access to a senior faculty position. The analogy is not exact, however. Unlike in the United States, where tenure commonly follows a relatively standardized probationary appointment as Assistant Professor and operates as an up-or-out decision, candidates in Spain could reach this stage from a range of prior positions, including temporary postdoctoral and teaching contracts as well as permanent teaching-track positions. Failure to obtain promotion therefore did not generally require immediate exit, and promotion

12. Spain has 48 public universities, not counting 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. In 2010, these universities employed approximately 100,000 academic staff, of whom about 9% were Full Professors, 28% Associate Professors, 17% tenured Assistant Professors (*Profesor Contratado Doctor*), and 12% non-tenured Assistant Professors (*Profesor Ayudante Doctor*). In addition, the remaining 35% held non-research positions including part-time teaching fellows and teaching-track staff not requiring a doctoral degree. Source: *Estadística de la Enseñanza Universitaria en España. Curso 2010/2011*, Instituto Nacional de Estadística.

did not necessarily mark the first attainment of employment security.

Faculty are also assigned to one of 188 nationally defined fields (*áreas de conocimiento*). Within universities, these fields usually correspond closely to the units responsible for hiring and promotion. We therefore define a department throughout the paper as a university–field cell. This definition captures the decision-relevant unit in most cases despite two institutional complications: some formal departments combine several fields (for example, management, accounting, and marketing) but organize hiring and promotion at the field level, while some large universities have more than one formal department within the same field (for example, multiple economics departments at Universidad Complutense).

2.2 The *Habilitación* System

Prior to 2002, Spanish universities had substantial autonomy over hiring and promotion, in a system characterized by strong localism in academic careers.[13] In 2002, the Spanish government introduced the *habilitación* system, a centralized promotion procedure aimed at increasing transparency and meritocracy. The reform separated access to senior academic ranks into two stages. First, candidates had to obtain a national qualification in their field and rank. Only after qualifying could they be appointed by a university to the corresponding position.[14]

In the *habilitación* system the number of national qualifications was capped. Universities first reported the number of positions they expected to open, and the Ministry then issued field-rank calls specifying the number of available qualifications. The cap meant that qualification was not simply a certification of minimum quality, but a binding step in the promotion process. The qualification stage was highly selective: only 11% of candidates qualified. In our data, more than 90% of qualified candidates were promoted within three years. In other words, promotion decisions were practically outsourced to the national qualification committees.

13. 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).

14. This system resembles centralized qualification procedures used in other countries, such as France's *agrégation* and Italy's *abilitazione scientifica nazionale*.

In each field, there were multiple rounds of public calls issued for Associate Professor and Full Professor qualifications. A single committee evaluated all candidates in a given field-rank call.

Candidates had twenty days to apply and application was voluntary. Only after the application deadline had passed did the Ministry draw the evaluation committee, so candidates could not condition their application decision on the realized committee composition.

Committee members were selected by lottery 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 to participate as an evaluator required meeting minimum research productivity thresholds set by the national research evaluation agency. Each evaluation committee consisted of seven members. Appointment to a committee was generally non-renounceable, and fewer than 2% of selected evaluators were replaced.[15] Participation in committees for earlier calls was not affecting the chances of being drawn again.

There were no conflict-of-interest rules implemented. Academic connections between evaluators and candidates did not require evaluators to be replaced, so some candidates were randomly assigned committees that included a strong connection, such as a coauthor, departmental colleague, or PhD adviser.

Candidates underwent a sequence of eliminatory evaluation stages, the structure of which depended on the academic rank. For candidates for Associate Professor, these stages included presentation of their CV teaching, a teaching demonstration, and a research seminar. Candidates for Full Professor had only two evaluation stages, without a teaching demonstration stage. Beyond the general structure of the evaluation process, committees had autonomy in setting up criteria. One third of committees concluded the evaluation process within one year, and most concluded within two years.

Following the final *habilitación* rounds (2006–2008), Spain transitioned to the current *acred-*

15. General abstention and recusal provisions remained applicable under 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, withdrawals owing to connections with candidates were exceedingly rare. For instance, of the 832 professors assigned to evaluate their own PhD students, only 22 opted out, a proportion similar to the overall withdrawal rate.

itación system. Although the new system retains a national qualification stage, universities regained greater discretion in promotions and hiring.

Promotion activity experienced two major slowdowns during our study period: first between 2004 and 2008, as the *habilitación* system created a bottleneck, and again between 2013 and 2016, when the government imposed a nationwide hiring freeze on public universities as part of austerity measures.[16]

Taken together, the institutional features of the *habilitación* system make it particularly well suited to studying the dynamic consequences of promotion. Since candidates applied before committees were drawn, the random assignment of evaluators generated quasi-random variation in exposure to connected evaluators and, consequently, in qualification outcomes. This variation was consequential: qualification constituted a binding national bottleneck, with a capped number of qualifications available. The replacement of the *habilitación* system in 2008 by a less restrictive and more decentralized procedure then allows us to trace how these exogenous promotion shocks affected departments once universities regained greater autonomy.

2.3 Doctoral Training

Doctoral education in Spain is organized within a nationally regulated, program-based framework.[17] Doctoral candidates are admitted to doctoral programs coordinated by Doctoral Schools, rather than being organized solely around bilateral adviser-student matches.

Each doctoral program is overseen by an Academic Committee that decides on admissions, assigns supervisors, approves research plans, and monitors progress through periodic evaluations. Lead supervisors must meet accreditation and research-activity requirements defined

16. The See El País (2013) and The Guardian (2013). Appendix Figure A1 documents the flow of promotions and the stock of academics in Spain. 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.

17. 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.

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 senior research-active faculty.

Doctoral registration does not necessarily imply full-time attachment to research. Among doctorate holders surveyed by the Spanish National Statistics Institute in 2009, only 34% reported having financed their doctoral studies primarily through public-administration grants, while about 40% worked to finance their doctorate (Instituto Nacional de Estadística, 2010). Moreover, completion appears far from automatic even among funded doctoral students. Matching state fellowship recipients to TESEO, the public repository of PhD dissertations in Spain, Buela-Casal et al. (2011) find that only about 40% of 2002–2003 fellows had a registered dissertation by 2009. A related study finds no statistically significant gender differences in completion among funded doctoral students (Bermúdez et al., 2011). These figures should not be interpreted as exact completion rates for all enrolled PhD students, but they indicate that effects on PhD graduates may operate through completion as well as recruitment.

3 Data

We construct a linked dataset combining national qualification exams, faculty careers, departmental composition, PhD outcomes, and research output. Building on Bagues et al. (2017), we use administrative records from the *habilitación* system on candidates, evaluators, committee assignments, and qualification outcomes. We link these records to promotions from the Spanish State Bulletin (*Boletín Oficial del Estado* – BOE), publication records from OpenAlex, Scopus, and Dialnet, and the near-universe of registered PhD dissertations from TESEO. The linked data allow us to follow faculty careers and departmental composition from 1985 to 2023, doctoral graduates from 1990 to 2017, and candidates exposed to the *habilitación* system between 2002 and 2008.[18]

18. Appendix G provides details on data sources, harmonization, validation, name disambiguation, and publication matching.

In this section, we first describe the *habilitación* records and the candidate-level treatment variation, then explain how we track individual-level and department-level outcomes of qualification decisions. We conclude by describing the measures of research specialization used to distinguish gender from the gender composition of research areas.

***Habilitación* records and connections.** The *habilitación* records cover 967 exams across 174 research fields: 465 for Associate Professor and 502 for Full Professor.[19] We observe 31,243 applications, with candidates submitting two applications on average, reflecting both repeated attempts across exam waves and simultaneous applications to closely related fields. For each exam, the records identify the pool of eligible evaluators and the committee members drawn by lottery; the average evaluator pool contained 114 researchers.

The linked data include candidates' and evaluators' gender, field, rank, university affiliation, qualification outcome, and research output. We use these records to construct candidate–committee connectedness. A candidate has a strong tie to an evaluator if the evaluator was a prior PhD adviser, coauthor, or departmental colleague. Approximately one third of applicants have at least one strong tie to a committee member: 35% in Full Professor exams and 29% in Associate Professor exams. Most ties arise from shared departmental affiliation, while coauthorship and adviser–advisee links are less common.

The qualification process was highly selective. Only about 11% of Full Professor candidates and 12% of Associate Professor candidates qualified. Applicants to Full Professor positions were substantially more senior, with 43 publications on average, compared with 21 among Associate Professor applicants, while average journal impact – as measured by the Article Influence Score (AIS) – was similar across ranks.[20] Women qualified at slightly lower rates than men at both ranks and had, on average, fewer publications and lower AIS at the time of application (Appendix Table A1). By contrast, women and men were similarly exposed to

19. 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.

20. 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.

connected evaluators, although tie intensity was slightly lower for women.

Table 1 shows that successful candidates are substantially more productive than the average applicant, whether or not they are connected to their evaluators. Connected successful candidates had somewhat weaker publication records than unconnected successful candidates, but this difference was modest relative to the larger productivity gap between successful candidates and the full applicant pool. This pattern is consistent with connections affecting outcomes near the qualification margin rather than allowing substantially weaker applicants to qualify.

Individual career outcomes. We follow candidates' subsequent academic careers using BOE promotion records, which report appointments to Associate and Full Professor positions in Spanish public universities. For each candidate, we measure whether and when they are promoted to the rank for which they applied, up to fifteen years after the qualification decision, and whether the promotion occurs in the same department where they were based at the time of the exam.

Qualification was closely linked to subsequent promotion. More than 90% of qualified candidates were promoted within two years, and among those promoted, 86% obtained their position in the same department where they were based at qualification (Appendix Figure A2). Some unsuccessful candidates later reapplied and qualified in subsequent rounds, generating gradual catch-up in promotion rates. This catch-up was more pronounced among Full Professor candidates, who typically already held tenured Associate Professor positions, than among Associate Professor candidates.

We also use the *habilitación* records to measure reapplication after an unsuccessful attempt. To characterize continued academic attachment, we measure academic activity as producing at least one journal article, book, or book chapter in a window around the relevant post-exam horizon.

Department-level faculty outcomes. For the department-level analysis, we define departments as university–field cells (for discussion of this see Section 2). Using BOE promotion records, we reconstruct annual department-level senior faculty stocks and promotion flows by rank and gender from 1985 to 2023.[21]

During the *habilitación* period, we identify 4,730 departments. The median professor works in a department with 11 tenured faculty members, roughly 20% of whom are Full Professors, but department size is highly skewed: about 1,000 researchers are the only tenured faculty member in their university–field cell, while only 20 departments have more than 50 tenured professors (Appendix Figure A3). Female representation also varies sharply by rank and field. Among Associate Professors, the share of women rose from 32% in 2000 to 43% in 2023; among Full Professors, it rose from 13% to 28% (Appendix Figure A4). The median STEM+M professor is in a department where 25% of tenured faculty are women, compared with 30% in SSH (Appendix Figure A5).

The causal analysis focuses on the 4,000 departments with at least one faculty member participating in a qualification exam. Female candidacies were less common than male candidacies, especially at the Full Professor level, and departments typically submitted one to two candidates per exam. Successful female qualifications were rare: 7% of departments qualified a female Full Professor candidate, compared with 20% qualifying a male candidate; at the Associate Professor level, the corresponding figures were 14% and 22%. Participating departments were larger than average, employing 11–13 tenured faculty compared with about 7 among all departments nationally. Over the ten years following a Full Professor exam, they promoted on average 1.7 men and 0.7 women to Full Professor; following Associate Professor exams, they promoted 2.4 men and 1.5 women to Associate Professor (Appendix Table A2).

21. We assign each researcher to the department in which they first obtain an Associate Professor position and assume they keep this affiliation fixed until age 65, unless subsequently hired by another public university. We validate the resulting department-level database using aggregated data obtained from the Spanish Ministry of Education through a data request. The Ministry data report departmental size and gender composition for 2011–2018, for departments with at least five female or male faculty members. Correlations between our reconstructed measures and the Ministry data exceed 90%, suggesting limited measurement error.

PhD graduates and early-career outcomes. We measure doctoral training using TESEO, the official registry of PhD dissertations defended in Spain. The data cover 203,542 theses defended in Spanish public universities between 1990 and 2017, after excluding approximately 8% with missing information on research field, graduation year, or adviser. The records include graduates' gender, university, field, graduation year, adviser, and dissertation topic. We assign graduates to the corresponding university–field department. The average time to completion is approximately five years.[22]

Women account for 47% of PhD graduates, while 21% of graduates have a female supervisor. Female students are more likely than male students to be advised by women (Panel A of Appendix Table A3).

Our main PhD outcomes are the cumulative number of female and male PhD graduates in a department after the qualification shock, the number who remain academically active five years after graduation, and the number promoted to Associate Professor within ten years of graduation. Around 50% of graduates keep publishing five years after graduation, consistent with survey evidence that academic careers absorb less than half of PhD graduates in Spain (Auriol et al., 2013); around 10% are promoted to Associate Professor within ten years.

Gender gaps in early-career 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 larger in STEM+M, where women are 7% less likely than men to remain academically active and 45% less likely to be promoted within ten years; in SSH, the gaps are more muted. (Panels B and C of Appendix Table A3). Women also produce fewer research outputs during and after the PhD, but with no difference in average AIS, suggesting differences in publication quantity rather than journal quality.

In the department-exam sample, departments graduate on average 13–15 PhD students of each gender over the ten years following a qualification exam.

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

Gender differences in research specialization. To distinguish the role of gender from the role of research specialization, we construct measures of the gender composition of research areas. 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, and aggregate these indices to the paper and author level.[23] An author's yearly index therefore captures the gender composition of the research topics in which they publish, rather than their own gender. For example, in 2000, *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 graduates, we use fields according the UNESCO Thesaurus classification reported at 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*). Appendix Figure A6 reports the Duncan segregation index, which measures the fraction of graduates who would need to be reallocated across topics for female and male topic distributions to coincide.[24] Segregation is sizable and persistent: at the six-digit level, the index declines from roughly 0.40 in the 1990s to just above 0.30 by 2015, while for broader two- and four-digit classifications it remains much more stable.[25]

4 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

23. 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 G.6 for more details.

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

25. Appendix Figure A7 shows the distribution of the topic gender composition index for professors, exam candidates, PhD graduates, and advisors. Women and men specialize in different research areas on average, but the overlap is substantial.

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.[26]

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.

The following subsections describe the identifying assumptions, specification choices, and interpretation of the key coefficients.

4.1 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

26. 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.

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.

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.[27] 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

27. 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.

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.

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 E1 and Appendix Figure E1), 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.

4.2 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.

5 Main Results

We organize the causal evidence in four parts. Sections 5.1 and 5.2 examine how qualification decisions affect candidates' careers, overall and by gender. Section 5.3 then analyzes whether exogenously promoting a woman changes the subsequent gender composition of her

department's faculty. Finally, Section 5.4 studies downstream effects on PhD cohorts and their academic outcomes. Throughout, we abbreviate percentage points as *p.p.* and standard errors as *s.e.*

5.1 The Effects of Qualification on Candidates' Careers

We begin by examining whether qualification decisions during the *habilitación* period affect long-run academic careers of researchers. The analysis exploits the random assignment evaluators to committees: candidates whose committees included more academic connections than expected were more likely to qualify. We use this variation to estimate the effect of passing the qualification threshold on subsequent career trajectories.

Table 2 reports first-stage estimates. Academic connections strongly increase qualification probabilities, consistent with Zinovyeva and Bagues (2015). 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 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.

Figure 1 plots 2SLS estimates of the effect of an exogenous qualification on promotion over a 15-year horizon, separately by rank. In the short run, effects are similar across ranks: one year after the committee decision, exogenously qualified candidates are about 75 p.p. (s.e. = 7) more likely to have been promoted. This falls below the more than 90% promotion rate of qualified candidates within two years reported in Section 3, reflecting some early catch-up by initially unsuccessful candidates.

The long-run effects diverge markedly between ranks. For Associate Professor candidates, the career advantage remains large and persistent: exogenously qualified candidates are still 54 p.p. (s.e. = 11) more likely to have been promoted after ten years and 40 p.p. (s.e. = 11) after fifteen years. For Full Professor candidates, by contrast, the effects decline steadily, become statistically insignificant after around six years, and are close to zero by the end

of the 15-year horizon. This pattern points to greater catch-up among unsuccessful Full Professor candidates, who typically already held tenured Associate Professor positions and could reapply from a stable academic position.

5.2 Gender Differences in the Career Effects of Qualification

We next estimate the career effects of qualification separately by gender. The first-stage relationship between committee connections and qualification remains strong for both women and men, with no statistically detectable difference across gender (Table 2). This allows us to compare how women's and men's subsequent careers respond to the same source of quasi-random variation in qualification.

At the Associate Professor level, qualification effects are more persistent for women than for men (Panels A and B of Figure 2). Ten years after the qualification decision, exogenously qualified women are 83 p.p. (s.e. = 17) more likely to have been promoted, compared with 38 p.p. (s.e. = 13) for men; the gender difference is marginally significant (p-value = 0.051). After 15 years, the corresponding effects remain at 57 p.p. (s.e.= 16) for women and 29 p.p. (s.e.= 12) for men. This slower convergence suggests that women are less likely to recover from an unfavorable committee draw, consistent with failed qualification having especially persistent consequences for their academic careers.

At the Full Professor level, the estimates for women are much less precise than those for men, reflecting the small number of female Full Professor candidates in the sample (Panels C and D of Figure 2). The point estimates suggest less persistent effects for women than for men, but the difference is not statistically significant.

To understand the gender differences at the Associate Professor level, Table 3 uses the same quasi-random variation to examine the consequences of narrowly failing the exam for reapplication in subsequent *habilitación* calls and academic activity. Among men who narrowly miss qualification, 42% reapply, while this response is 28 p.p. smaller for women. Failure is also associated with lower subsequent academic activity, with larger effects for women: fifteen years after the qualification decision, the estimated effect of failure is -8 p.p.

for men and an additional -30 p.p. for women. These patterns suggest that women who narrowly fail are less likely to re-enter the qualification process and less likely to remain active in research, helping explain the more persistent promotion effects for women.

5.3 The Effects of Female Promotions on Departmental Faculty Composition

We next ask whether individual qualification shocks translate into persistent changes in departmental faculty composition. The department-level design aggregates the same lottery-induced variation to the department–exam level: departments whose candidates draw more connections than expected experience more qualifications, and we use this variation to estimate the effect of an additional qualified candidate on subsequent promotion flows. 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.

Table 4 reports the department-level first stage. An additional realized connection between a department's candidates and the evaluation committee increases the number of qualified candidates in the department by 0.15 ($s.e. = 0.01$). In line with our individual evidence, each qualification translates into roughly one additional promotion at the corresponding rank in the short run (Appendix Figure F1). Over time, however, the effects diverge by rank: Associate Professor qualifications generate increasingly persistent department-level effects, while Full Professor effects fade. This likely reflects the fact that unsuccessful Full Professor candidates typically remain in the department as tenured Associate Professors and eventually manage to catch up. We therefore focus the remaining department-level analysis on Associate Professor qualifications.

Before examining gender-specific department-level effects, we verify that the first stage is similar for female and male candidates: departments whose female or male candidates draw more favorable committees experience comparable increases in the number of qualifications (Table 4). Figure 3 then decomposes Associate Professor qualification shocks by the gender of the qualifying candidate. Exogenously qualifying a female candidate increases the number of female Associate Professor promotions in her department by about 0.6 ($s.e. = 0.2$) within

one year, mechanically reflecting the promotion of the qualifying candidate. The effect then rises over time, reaching approximately 1.5 ($s.e. = 0.6$) additional female promotions after thirteen years.

The increase beyond the short-run direct effect suggests that qualifying a marginal woman may also help sustain the department's subsequent female promotion pipeline. This need not imply mechanical department expansion. Rather, the initial promotion may affect whether other women in the department remain active in research, seek later qualification or accreditation, and progress to promotion.

These effects are specific to female qualifications at the Associate Professor margin. Female qualification shocks have no significant effect on male Associate Professor promotions at any horizon, suggesting no offsetting reduction in male promotions. Conversely, male qualification shocks generate no significant long-run effects on either female or male Associate Professor promotions.[28]

Taken together, these results show that exogenously promoting a woman to Associate Professor generates a persistent increase in female representation within the department over the following fifteen years. Although we cannot isolate faculty-level spillovers, the estimates confirm a sustained shift in departmental gender composition, which we next exploit to study downstream effects on doctoral students.

5.4 The Downstream Effects of Female Promotions on PhD Cohorts

We next examine whether female Associate Professor promotion shocks has downstream effects on PhD cohorts and their academic outcomes. These estimates capture the combined department-level effect of the initial female qualification and any later changes in faculty composition that it induces.

Figure 4 traces cumulative effects on PhD graduates. Departments that exogenously promote

28. Appendix Figure F2 shows a positive but less precise effect on the stock of tenured women, consistent with stock measures being affected by additional margins such as mobility, attrition, and retirement. Gender-specific results for Full Professor qualifications, reported in Appendix Figures F3 and F4 show no significant long-run effects and are imprecise for female candidates.

a woman train substantially more female PhD graduates over time (Panel A). The effect begins to emerge three to four years after qualification and becomes statistically significant around year five, broadly consistent with the typical duration of doctoral training. Since we observe PhD completions rather than enrollment, the effect may reflect higher recruitment, higher completion, or both. After ten years, departments have graduated about six additional female PhD students. There is no significant effect on male PhD graduates (Panel B), suggesting that the increase does not come from within-department substitution away from male students.

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.[29] 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.[30] These effects are specific to female PhD graduates: female qualification shocks have no significant effect on male graduates' academic activity or later promotion.

Male Associate Professor qualification shocks do not generate comparable downstream gains. If anything, they are associated with a modest decline in female PhD outcomes (see Panel C in Figures 5 and 6), although these estimates are only marginally significant.

We also assess whether the increase in female PhD graduates reflects changes in supervision patterns or broader departmental changes. Figure 7 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

29. The average department graduates approximately 13 female PhD students over the decade following the exam with 54%, or about seven, of them remaining active in research (Appendix Table A2). Following a female qualification shock, the same department would graduate approximately 13+6=19 female PhD students, of whom about 7+4=11 would remain academically active, raising the retention rate to 58%.

30. Of the 13 female PhD students about half obtains promotion to Associate Professor within ten years of graduation, or 4% (Appendix Table A2). Following the exogenous female promotion, out of the 19 students, one and a half students become Associate Professors within ten years of graduation, raising the promotion rate to 8%.

attractive or supportive for female doctoral students more broadly.

Overall, the evidence suggests that promoting a marginal woman affects not only the gender composition of senior faculty, but also the subsequent female PhD pipeline: more women complete PhDs, more remain active in research, and more eventually reach Associate Professor.

5.5 Heterogeneity by Department Composition and Field

We next examine whether the downstream effects of female Associate Professor qualifications differ across departments. We consider two dimensions: baseline female representation among senior faculty and broad disciplinary group. These dimensions are related but not equivalent: female representation tends is lower on average in STEM+M, but many SSH departments also have few senior women (Appendix Figure A5). The representation split asks whether effects depend on the department's initial gender composition, while the field split captures broader differences in doctoral training, research organization, and career paths. Because the subgroups are smaller, we interpret these estimates as suggestive.

Panel A of Table 5 splits departments by the pre-exam share of women among Associate and Full Professors: no senior women, women present but in the minority, and women in the majority. The largest effects appear in departments where women are present but remain a minority. In this group (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, estimates 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). Although the subgroup coefficients are not statistically different from one another, the pattern is consistent with the idea that female promotions have the largest scope to matter where there is already some female presence, but substantial room for further progress.

Panel B of Table 5 reports estimates separately for STEM+M and SSH departments. Effects are positive in both groups, and we cannot reject equality across fields. The point estimates

suggest somewhat larger effects on female faculty promotions in STEM+M, while effects on female PhD graduates and their subsequent academic careers are somewhat more precise in SSH. Overall, the analysis suggests that the main effects are not confined to a single broad field group.

5.6 Gender and Research Orientation

The effects documented above may operate not only through the gender of the promoted candidate, but also through their research areas. Research areas differ substantially in their gender composition, and promoting candidates in female-intensive areas could strengthen a department's female pipeline even when the promoted candidate is not a woman. We therefore examine whether qualifications in female-intensive research areas generate downstream effects comparable to those of female qualifications.

Using the feminisation index described in Section 3, we classify candidates as working in *female-intensive* research areas depending if their index lies above the mean within their field at the time of the exam. This classification overlaps with gender but is not a proxy for it: the average feminisation index is 30% for female candidates and 26% for male candidates, and 37% of candidates in female-intensive areas are women, compared with 28% in male-intensive areas. We apply the same department-level IV strategy to estimate the effect of exogenously qualifying a candidate in a female-intensive area.

Figure 8 shows that qualifications in female-intensive areas generate persistent effects on female faculty promotions. Within six years, departments have promoted 0.9 additional women (s.e. = 0.34). After fifteen years, the effect reaches 1.6 additional women (*s.e.* = 0.58), while the corresponding effect on male promotions is small and imprecise (Appendix Table F1). These effects are similar in magnitude to those generated by qualifying a female candidate directly, suggesting that research orientation is an important complementary channel for shifting faculty gender composition.[31]

31. Appendix Table F1 also shows no clear effects of qualifying candidates in male-intensive research areas on female or male promotions or PhD outcomes.

The effects on PhD outcomes are positive but smaller and less precise. Over ten years, qualifying a candidate in a female-intensive area generates approximately four additional female PhD graduates (Panel C, $\beta = 3.7$, $s.e. = 2.0$), around two additional female graduates who remain research-active (Panel E, $\beta = 1.8$, $s.e. = 1.1$), and about 0.5 additional female graduates who are promoted to Associate Professor within ten years of graduation (Panel G, $\beta = 0.49$, $s.e. = 0.25$). The effects on male PhD outcomes are limited throughout (Appendix Table F1). Since the faculty effects of research orientation accumulate gradually, the smaller PhD effects over a ten-year horizon should be interpreted cautiously.

Finally, we estimate a horse-race specification including both the number of qualified female candidates and the number of qualified candidates working in female-intensive research areas. For faculty promotions, the coefficient on female-intensive qualifications remains positive and statistically significant: conditional on candidate gender, qualifying a candidate in a female-intensive area increases female Associate Professor promotions by 1.2 ($s.e. = 0.64$) over fifteen years (Appendix Table F2). The coefficient on female qualifications is close to one but imprecisely estimated. For PhD outcomes, however, we observe the opposite pattern: the effect of female qualifications on female PhD graduates remains large and stable after controlling for research orientation, while the coefficient on female-intensive qualifications is smaller and imprecise (Appendix Table F2).

Taken together, the results suggest that research orientation is an important complementary channel, especially for faculty promotions. Expanding female-intensive areas appears to strengthen the female promotion pipeline, but it does not fully replicate the downstream effects of promoting women on PhD cohorts. This pattern is consistent with the interpretation that female faculty matter for PhD outcomes not only through the research areas they represent, but also through gender-specific channels such as mentoring, role models, networks, or departmental climate.

6 Conclusions

Women's persistent underrepresentation in academia has motivated policies aimed at increasing female representation at senior career stages. Yet credible evidence on whether these interventions generate broader effects along the academic pipeline remains limited. This paper studies this question using quasi-random variation in promotion outcomes generated by Spain's centralized qualification system. By linking national qualification exams to faculty careers, department-level promotions, PhD dissertations, and subsequent academic outcomes, we trace how shocks at a key career bottleneck propagate over time.

Our first result is that failing to qualify for Associate Professor has large and persistent career consequences, especially for women. Female candidates who narrowly miss qualification are much less likely to be promoted even ten to fifteen years later, and this persistence is substantially stronger than for male candidates. The evidence suggests that this is not simply delayed promotion: after an unfavorable qualification shock, women are less likely to reapply and less likely to remain active in research. A key implication is that gender gaps in senior academia are shaped not only by who is promoted, but also by who is lost after setbacks at key promotion margins.

Our second result is that promoting a marginal woman affects her department beyond her own career. An exogenous female qualification to Associate Professor generates a persistent increase in subsequent female Associate Professor promotions, with effects rising beyond the short-run direct promotion of the qualifying candidate. These effects are strongest in departments where women are present but still underrepresented, suggesting that female promotions may be most consequential where there is already some female presence but substantial scope for change. We find little evidence that these gains come at the expense of male promotions within the same department.

The effects also extend to the PhD pipeline. Departments in which a marginal woman qualifies train more female PhD graduates, more female graduates who remain active in research, and more female graduates who later reach Associate Professor. These effects are

not confined to students supervised by women, pointing to broader department-level changes rather than only direct advising relationships. Nor do the estimates appear to reflect simple within-department displacement: female qualification shocks do not significantly reduce male promotions, male PhD graduates, or male graduates' subsequent academic outcomes.

The evidence points to several possible mechanisms. At the individual level, unfavorable qualification shocks are especially persistent for women because they are less likely to reapply and less likely to remain active in research after failure. At the department level, the effect of a female qualification extends beyond the direct promotion of the qualifying candidate, suggesting that the initial shock helps sustain the subsequent female promotion pipeline. Research orientation is also an important complementary channel: promoting candidates in female-intensive research areas increases later female faculty promotions. However, these research-orientation effects do not fully replicate the effects of promoting women on PhD outcomes, suggesting that female faculty matter for doctoral students not only through the topics they represent, but also through gender-specific channels such as mentoring, role models, networks, or departmental climate.

Cross-department displacement is harder to rule out directly. However, a pure reallocation story would imply substantial sorting of female students or researchers across departments in response to local promotion shocks, which seems less likely in a setting with limited faculty and doctoral-student mobility. Moreover, the effects extend beyond PhD completion to later academic activity and promotion, suggesting changes in subsequent trajectories rather than only where students complete the PhD.

These findings have implications for the design of gender-equity policies in academia. The compliers in our setting are productive researchers at the margin of qualification, so our estimates speak to policies that reduce promotion frictions for strong female candidates rather than to mechanical increases in representation or lower standards. Policies that prevent capable women from being derailed by arbitrary or temporary setbacks — for example by reducing idiosyncratic barriers in promotion decisions, expanding opportunities to reapply, or providing support after unsuccessful evaluations — may have effects that extend well beyond the careers of the directly affected candidates.

More broadly, the results highlight the importance of considering the downstream consequences of promotion decisions at key academic career bottlenecks. Promoting a marginal woman at a pivotal career stage can help sustain the female promotion pipeline, increase the number of female PhD graduates, and improve their subsequent academic trajectories. Understanding these longer-run effects is essential for evaluating policies that aim not only to change who holds senior positions today, but also to shape who enters, remains in, and advances through academia in the future.

References

Adams-Prassl, **Abi**, **Kristiina Huttunen**, **Emily Nix and Ning Zhang**. 2024. "Violence against Women at Work". *The Quarterly Journal of Economics* 139 (2), pp. 937–991.

Antecol, **Heather**, **Kelly Bedard and Jenna Stearns**. 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" *American Economic Review* 108 (9), pp. 2420–41.

Auriol, **Emmanuelle**, **Guido Friebel and Sascha Wilhelm**. 2020. "Women in European Economics". *Women in Economics*. Ed. by **Shelly Lundberg**. London: CEPR Press, pp. 26–30.

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).

Bermúdez, **Mª Paz**, **Alejandro Guillén-Riquelme**, **Almudena Gomez-Garcia**, **Raul Quevedo-Blasco**, **Juan-Carlos Sierra and Gualberto Buela-Casal**. 2011. "Analisis Del Rendimiento En El Doctorado En Funcion Del Sexo". *Educación XX1* 14, pp. 17–33.

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.

Buela-Casal, **Gualberto**, **Alejandro Guillén-Riquelme**, **Ma Paz Bermúdez and Juan-Carlos Sierra**. 2011. "Análisis Del Rendimiento En El Doctorado En Función de Las Becas de Formación de Profesorado Universitario y de Formación de Personal Investigador". *Culture and Education* 23 (2), pp. 273–283.

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

Buser, **Thomas and Huaiping Yuan**. 2019. "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad". *American Economic Journal: Applied Economics* 11 (3), pp. 225–252.

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.

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.

Instituto Nacional de Estadística. 2010. *Survey on Human Resources in Science and Technology. Year 2009. Press Release.* https://www.ine.es/en/prensa/np621_en.pdf.

Janys, **Lena**. 2024. "Testing the Presence of Implicit Hiring Quotas with Application to German Universities". *The Review of Economics and Statistics* 106 (3), pp. 627–637.

Kang, **Le**, **Ziteng Lei**, **Yang Song and Peng Zhang**. 2024. "Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College

Entrance Exam Retakes". *Journal of Political Economy Microeconomics* 2 (2), pp. 355–397.

Koffi, **Marlène**. 2021. "Gendered Citations at Top Economic Journals". *AEA Papers and Proceedings* 111, pp. 60–64.

— 2025. "Innovative Ideas and Gender (In)Equality". *American Economic Review* 115 (7), pp. 2207–2236.

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.

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.

Sarsons, **Heather**, **Klarita Gërxhani**, **Ernesto Reuben and Arthur Schram**. 2021. "Gender Differences in Recognition for Group Work". *Journal of Political Economy* 129 (1), pp. 101–147.

Tesar, **Linda**. 2025. "Report of the Committee on the Status of Women in the Economics Profession". *AEA Papers and Proceedings* 115, pp. 918–943.

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.

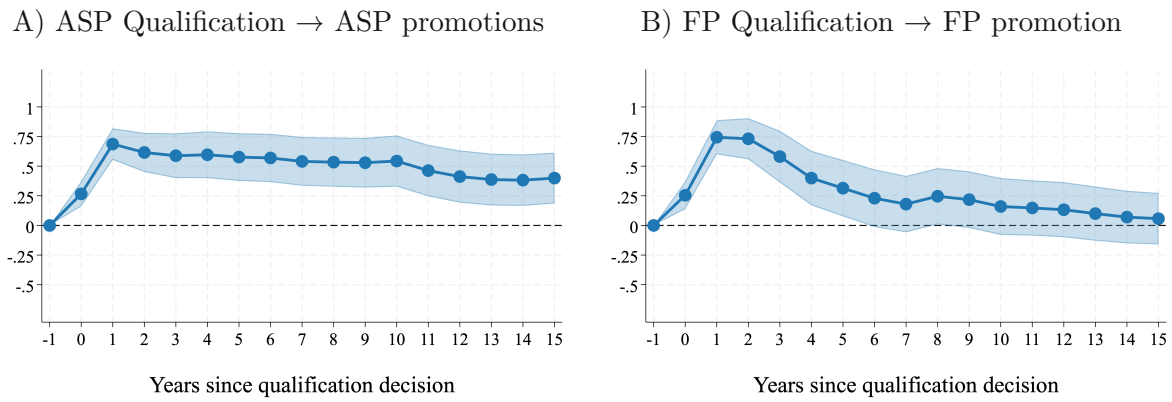
Wasserman, **Melanie**. 2023. "Gender Differences in Politician Persistence". *The Review of Economics and Statistics* 105 (2), pp. 275–291.

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.

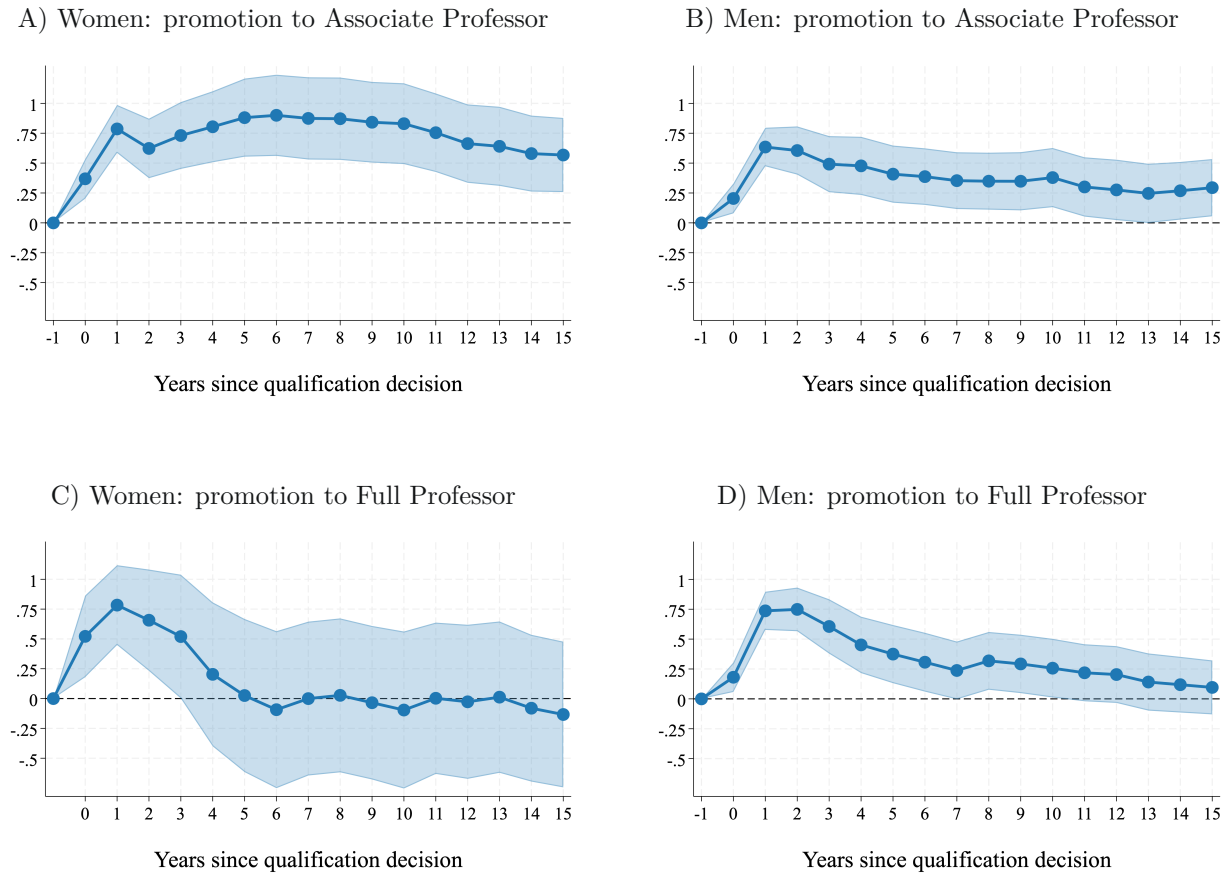
Figures

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



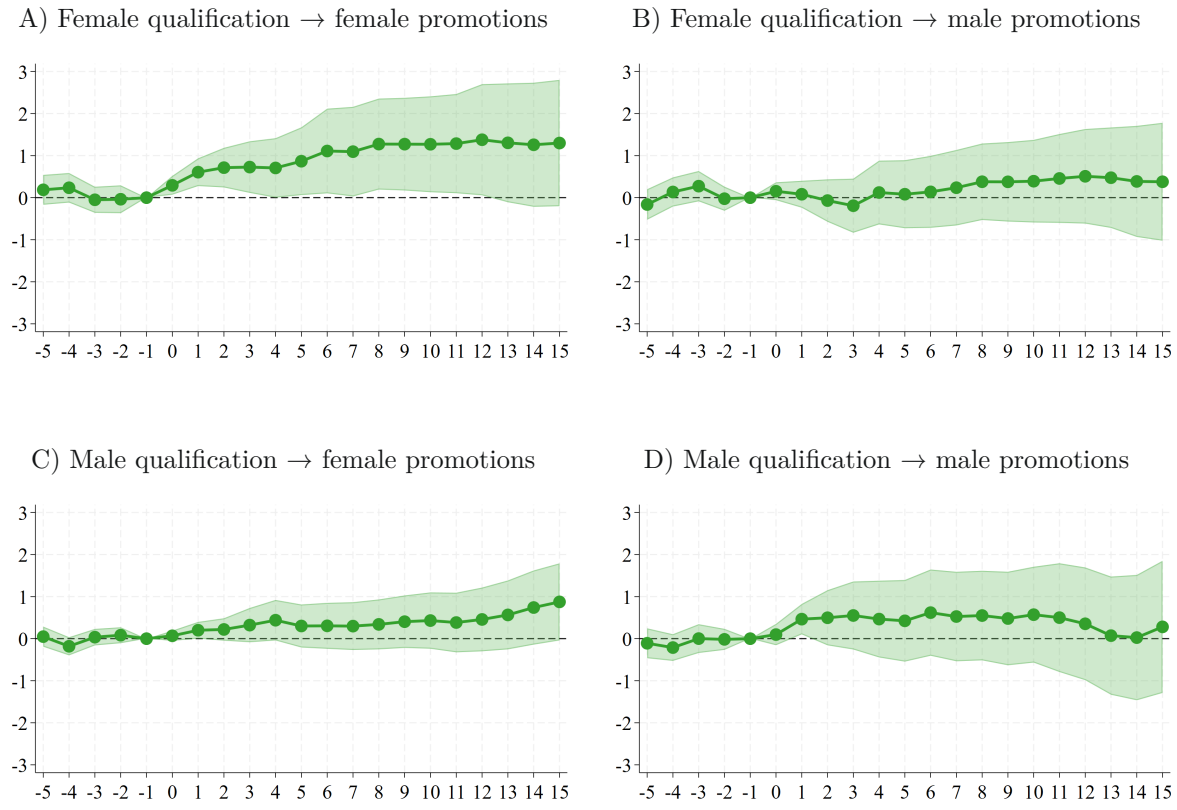
Notes: The figure reports 2SLS estimates of coefficient β_1 from Equation 2, capturing the impact of exogenous qualification in a given rank on the probability of promotion to that same rank in the years following the exam. Panel A reports effects on promotion to Associate Professor, while Panel B reports 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



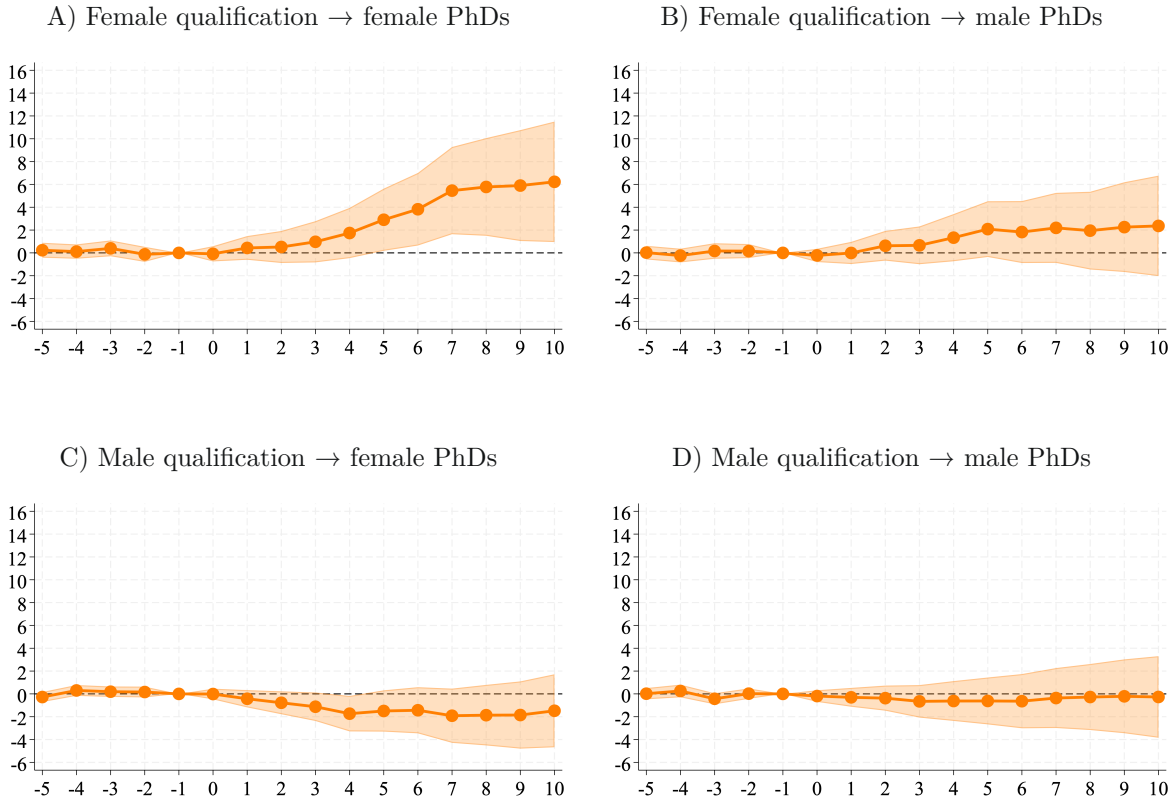
Notes: The figure reports 2SLS estimates of coefficient β_1 from Equation 2, capturing the impact of exogenous qualification in a given rank on the probability of promotion to that same rank in the years following the exam. Panels A and B report effects on promotion to Associate Professor for women and 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*



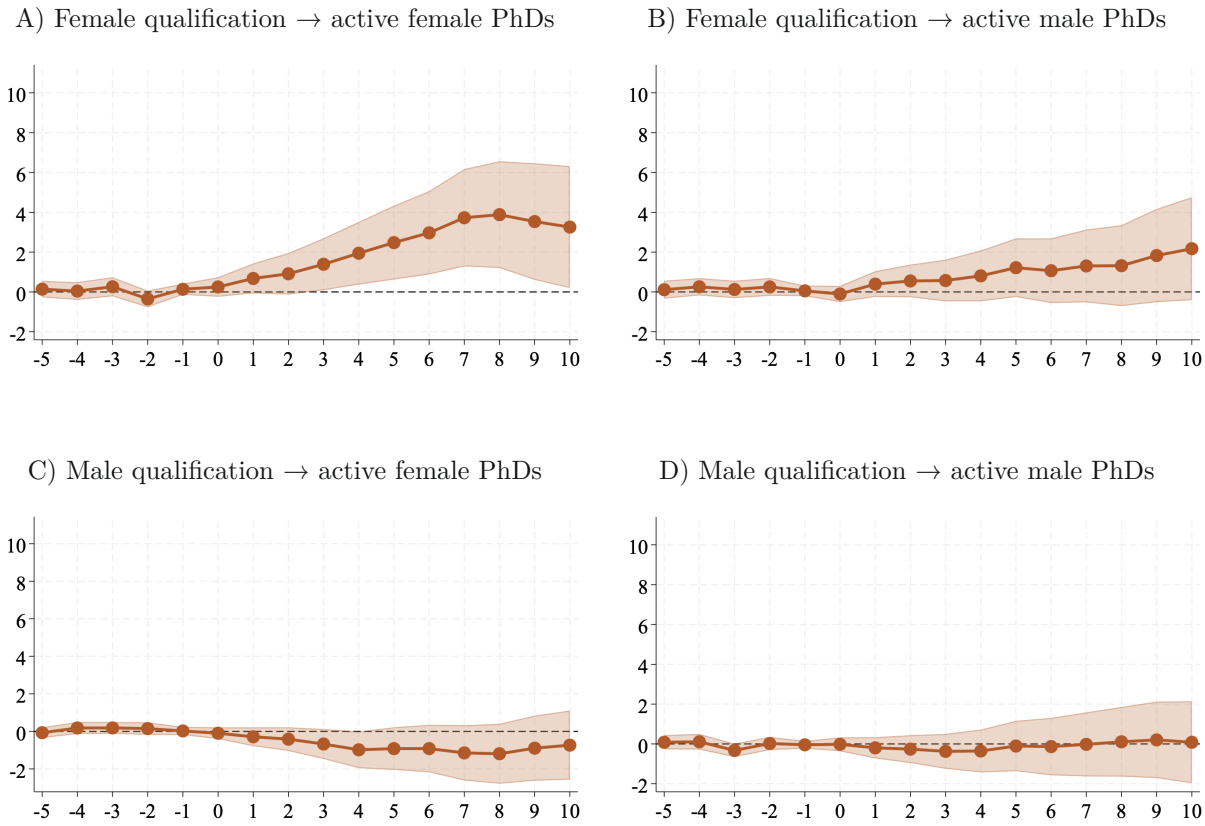
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6). Panels A and B show the effect of an exogenous female qualification to *Associate Professor* on the number of female and male promotions to *Associate Professor*, respectively. Panels C and D show the corresponding 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 4.2. 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



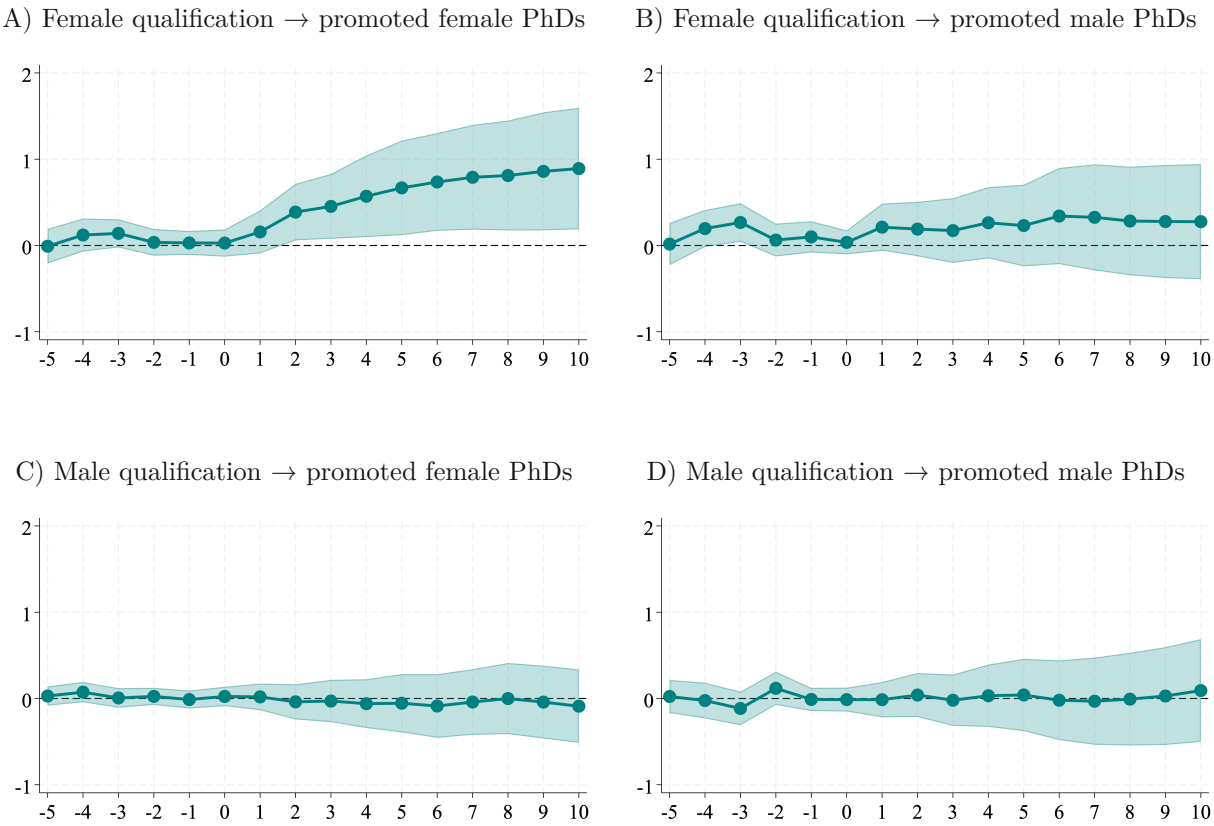
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from Equation (6), capturing the impact of exogenously qualifying a woman or a man to Associate Professor on the number of female and male PhD graduates. Outcomes are measured relative to the year of qualification and are cumulative 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 4.2. 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 4.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

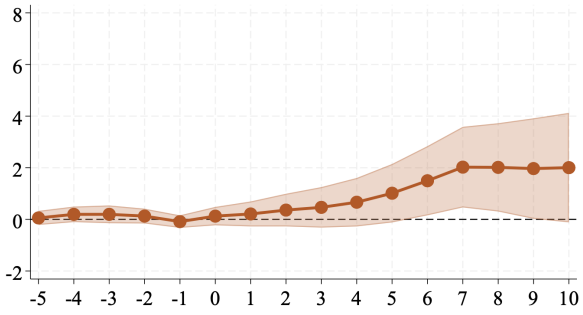
Figure 6: Department-level effects of female and male qualifications to *Associate Professor* on the number of PhD graduates who get promoted to Associate Professor within 10 years of graduation



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 4.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure 7: 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

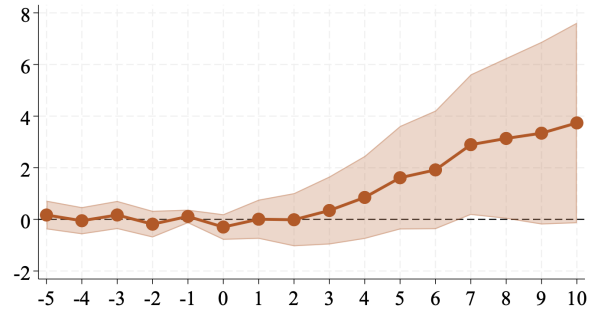
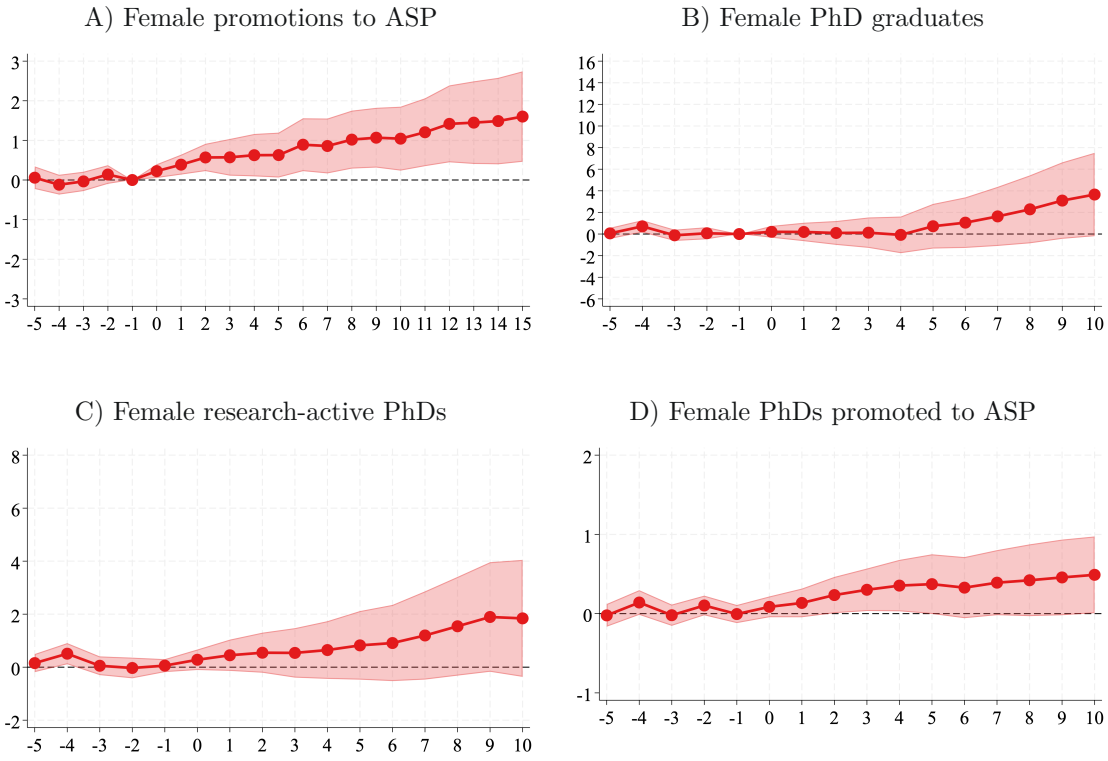


Figure 8: Department-level effects of *ASP* qualifications in female-intensive research areas



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 4.2. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Tables

Table 1: 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 2: First-stage estimates, individual level

	Full Professor Exams				Associate Professor Exams			
Outcome: Qualified	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All candidates		Men	Women	All candidates		Men	Women
Strong Connections	0.052***	0.052***	0.055***	0.040***	0.064***	0.064***	0.068***	0.061***
	(0.005)	(0.005)	(0.006)	(0.009)	(0.006)	(0.006)	(0.007)	(0.008)
Productivity controls:								
- Journal Articles		0.025***	0.023***	0.027***		0.037***	0.036***	0.040***
		(0.005)	(0.005)	(0.009)		(0.004)	(0.005)	(0.007)
- Total AIS		0.025***	0.029***	0.013*		0.008**	0.004	0.019***
		(0.004)	(0.005)	(0.007)		(0.004)	(0.005)	(0.007)
- Book Chapters		0.014***	0.012***	0.024***		0.014***	0.011***	0.019***
		(0.003)	(0.004)	(0.006)		(0.003)	(0.004)	(0.005)
- Books		0.001	-0.001	0.008		0.001	-0.004	0.007
		(0.003)	(0.003)	(0.007)		(0.003)	(0.003)	(0.005)
Observations	13439	13439	9802	3594	17762	17762	10687	7069
KP rk Wald F-stat	120.147	126.545	96.741	19.264	129.455	131.548	97.152	52.409

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 3: Impact of failing qualification on re-application and academic activity by gender

	Reapplication	Academic retention after		
	(1)	(2)	(3)	(4)
	Within habilitación system	5 years	10 years	15 years
Failed	0.418***	-0.134	-0.153*	-0.076
	(0.075)	(0.087)	(0.092)	(0.099)
Failed × Female	-0.279**	-0.148	-0.207	-0.298*
	(0.120)	(0.136)	(0.157)	(0.153)
Observations	17762	17762	17762	17762

Notes: The table reports 2SLS estimates of the effect of failing the national qualification exam on subsequent re-application in the *habilitaci'on* system and on academic activity five, ten, and fifteen years after the qualification decision. Re-application is an indicator equal to one if the candidate applies again in a subsequent call. Academic activity at horizon (t) is an indicator equal to one if the candidate has at least one publication, book, or book chapter in a three-year window centered on year (t) after the qualification decision. All specifications control for baseline productivity, expected committee connection intensity, exam fixed effects, and year fixed effects. Standard errors, clustered at the exam level, are in parentheses.

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

Table 4: 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 5.6. *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 4.2. 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

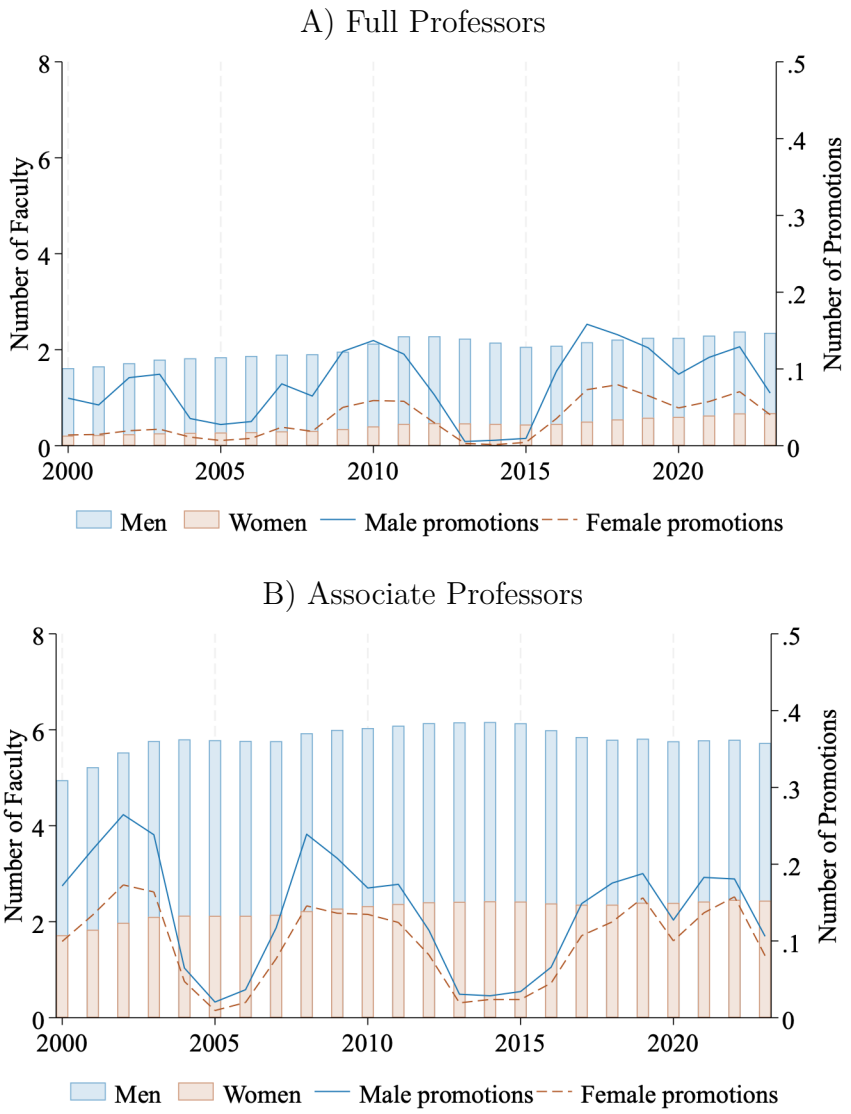
Table 5: Impact of female qualification by baseline female representation and field group

	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 ($\leq$ 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)
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)

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 4.2. All specifications include controls listed in Section 4.2. 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 Sciences, Technology, Engineering and Medicine.
*** p $<$ .01, ** p $<$ .05, * p $<$ .1

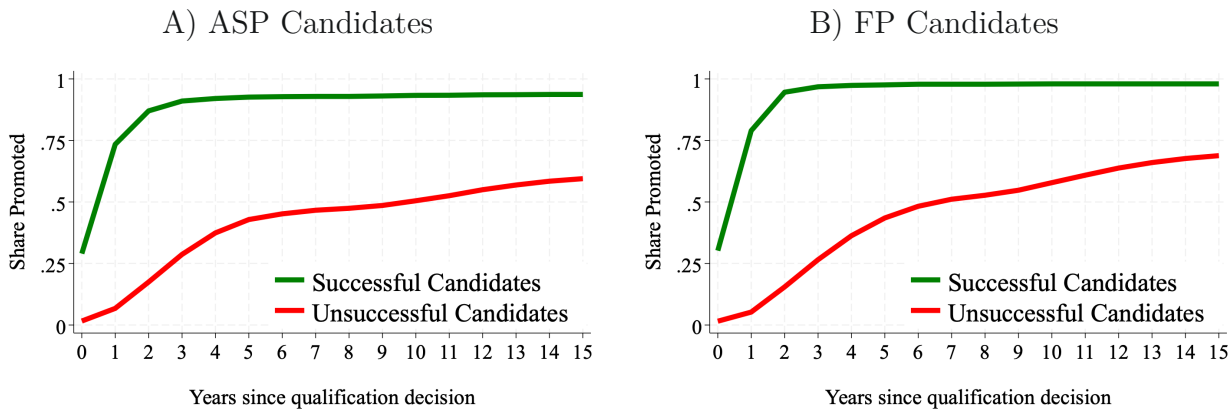
Appendix A — Descriptive statistics

Figure A1: Annual promotion flows and stocks of Associate and Full Professors



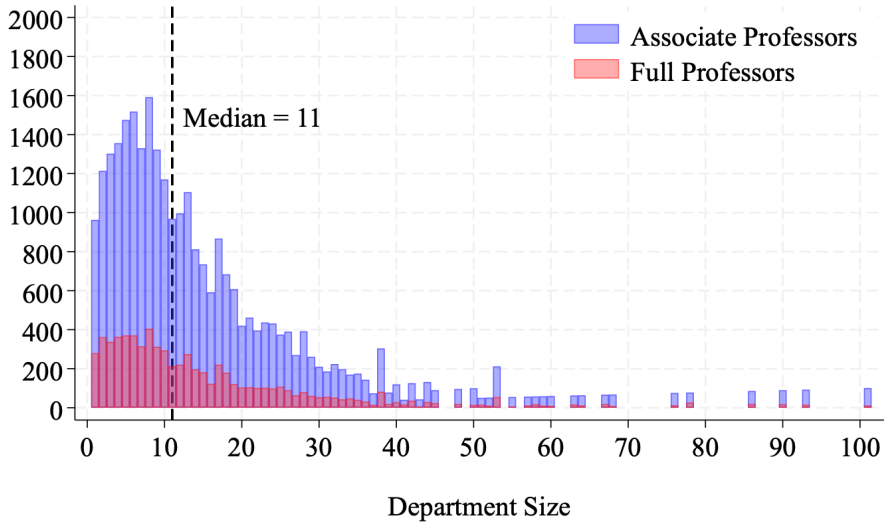
Notes: The figure shows annual promotions (lines) and the corresponding stocks (bars) of Full Professors (Panel A) and Associate Professors (Panel B) in Spanish public universities over time. Our own calculations using data from the Spanish State Bulletin (BOE).

Figure A2: Promotion rates by qualification outcome and years since exam



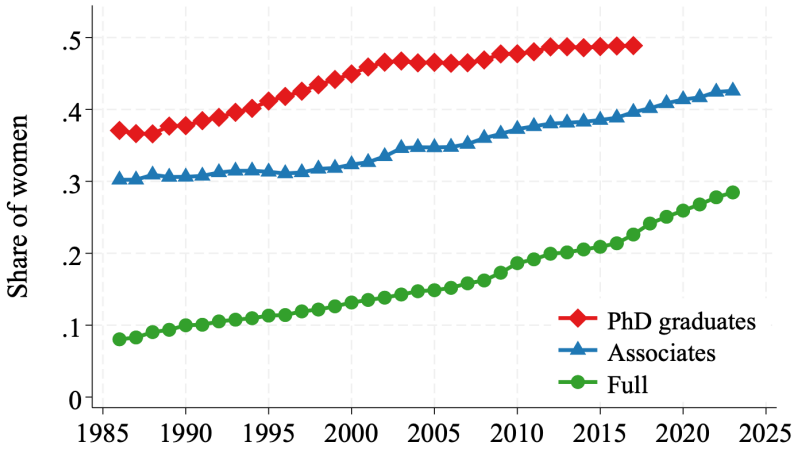
Notes: The figure plots the cumulative share of applicants who have been promoted, separately by qualification outcome, for candidates applying to Associate Professor positions (Panel A) and to Full Professor positions (Panel B). Outcomes are measured relative to the year of the qualification exam.

Figure A3: Distribution of tenured faculty over department size



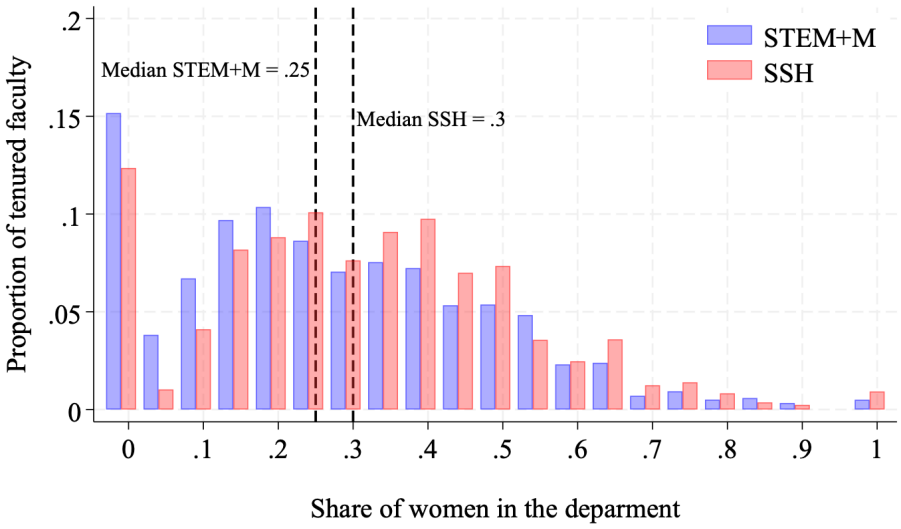
Notes: The figure displays the distribution of tenured faculty by department size in 2000. The histogram separately reports Associate Professors and Full Professors, with the vertical dashed line indicating the median department size (11 faculty members). The underlying data are drawn from the Spanish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

Figure A4: Share of women by academic role over time



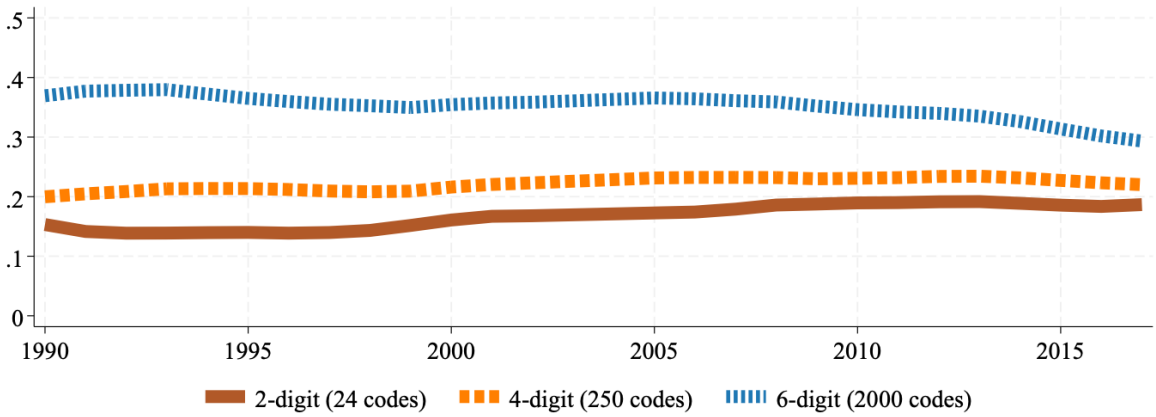
Notes: The figure displays trends in the share of women by academic role. The series report the fraction of women among PhD graduates (1985–2017) and among Associate and Full Professors (1985–2023), by year. Data are from PhD theses, appointment records of faculty eligible to serve on qualification exams,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

Figure A5: Distribution of tenured faculty over the share of women in the department



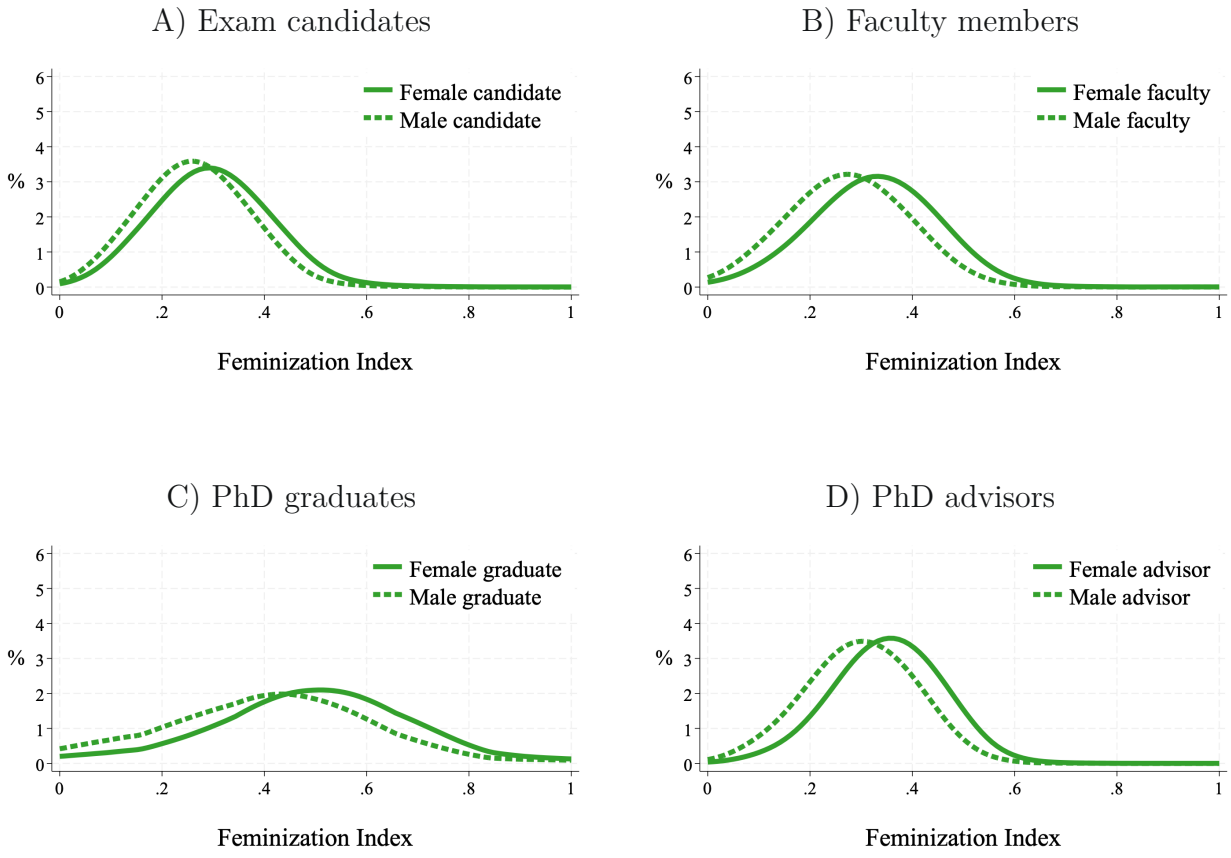
Notes: The figure displays the distribution of tenured faculty by the departmental share of women in 2000. The histogram separately reports STEM+Medicine and Social Sciences and Humanities (SSH) departments. The vertical dashed lines indicate the median departmental share of women in each broad field (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 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: 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 G.6 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.

Table A1: Characteristics of candidates in qualification exams by gender

	Overall	Men	Women	Women–Men	SE
	(1)	(2)	(3)	(4)	(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 A2: Characteristics of departments participating in qualification exams

	Departments with applicants to Full Professor		Departments with applicants to Associate Professor	
	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)
Panel A. Department-exam level characteristics				
Applicants to corresponding level				
Female	0.56	0.87	0.93	1.25
Male	1.50	1.44	1.42	1.77
Qualified to corresponding level				
Female	0.05	0.23	0.11	0.36
Male	0.17	0.42	0.18	0.46
Stock of Associate Professors 1 year prior to the exam				
Female	3.50	4.35	2.92	4.13
Male	6.08	6.88	5.39	7.06
Stock of Full Professors 1 year prior to the exam				
Female	0.42	0.85	0.37	0.77
Male	2.54	3.13	2.29	3.20
Panel B. Cumulative outcomes 10 years post-exam				
Promotions to Full Professor				
Female	0.71	1.24	0.48	1.00
Male	1.74	2.08	1.33	1.94
Promotions to Associate Professor				
Female	1.36	2.10	1.49	2.17
Male	1.98	3.16	2.36	3.51
PhD graduates				
Female	14.12	16.24	12.82	15.34
Male	14.92	16.05	14.26	15.84
PhD graduates staying active in academia				
Female	7.56	8.84	6.92	8.32
Male	7.96	8.81	7.63	8.71
PhD graduates promoted to Associate Professor				
Female	0.46	0.94	0.53	1.05
Male	0.79	1.63	0.91	1.79
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.

Table A3: 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	0.03	-0.03	-0.05***	0.01
Total Article Influence Score	0.81	0.02	-0.02	-0.05***	0.01
Average Article Influence Score	1.13	0.00	-0.00	-0.01	0.01
Accumulated research 1-5 years after graduation (conditional on staying in academia)					
Papers	1.76	0.40	0.33	-0.07***	0.02
Papers in Web of Science	1.19	0.36	0.32	-0.04*	0.02
Total Article Influence Score	1.47	0.33	0.29	-0.04*	0.02
Average Article Influence Score	1.18	0.03	0.01	-0.02	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	-0.02	-0.04***	0.01
Papers in Web of Science	0.09	0.01	-0.01	-0.01	0.01
Total Article Influence Score	0.08	0.00	-0.00	-0.01	0.01
Average Article Influence Score	0.79	0.00	-0.00	-0.01	0.04
Accumulated research 1-5 years after graduation (conditional on staying in academia)					
Papers	0.85	0.37	0.25	-0.12***	0.03
Papers in Web of Science	0.17	0.29	0.19	-0.10**	0.05
Total Article Influence Score	0.15	0.25	0.19	-0.06	0.05
Average Article Influence Score	0.72	-0.06	0.10	0.16*	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

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.[32] 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

32. 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.[33] 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:

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

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,

33. 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 check

This appendix reports diagnostic evidence supporting the identification strategy by presenting balance tables at the individual and department level.

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 (including exams both Associate and Full Professor 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 standardized 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 (including exams both Associate and Full Professor level) and department-level standardized 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

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 of our individual level results 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 candidates whose qualification status is affected by random committee assignment. We estimate the shares and observable characteristics of always-takers, compliers, and never-takers, and compare OLS and 2SLS estimates to clarify the interpretation of the IV results.

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 this exercise, 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, 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.[34] 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 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

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

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 E1 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 E1 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.[35]

35. 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

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 E2 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.

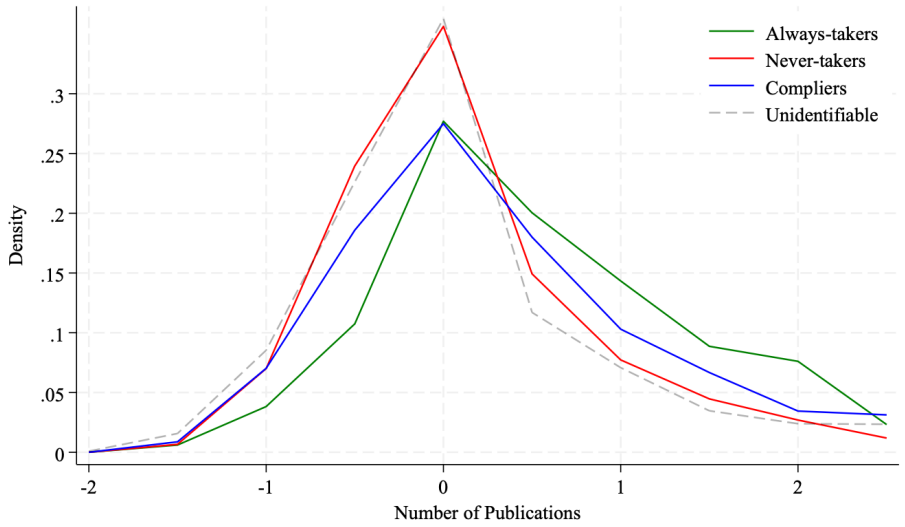
Appendix Figure E3 plots the distribution of subpopulations along this predicted ranking. Always-takers and compliers are concentrated among higher-ranked candidates, whereas never-takers are disproportionately represented toward lower ranks. Above the qualification margin, always-takers exhibit the highest density, followed by compliers. These patterns are similar for both men and women (Appendix Figure E4).

OLS vs. 2SLS. Appendix Table E2 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.

assumption. In this analysis we prioritize the internal validity of the estimates and tolerate a larger share of unidentifiable candidates.

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.

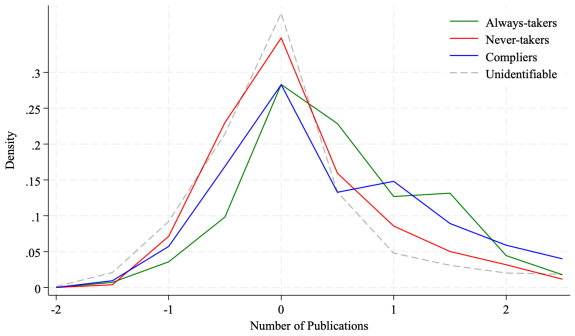
Figure E1: Productivity distribution by subpopulation



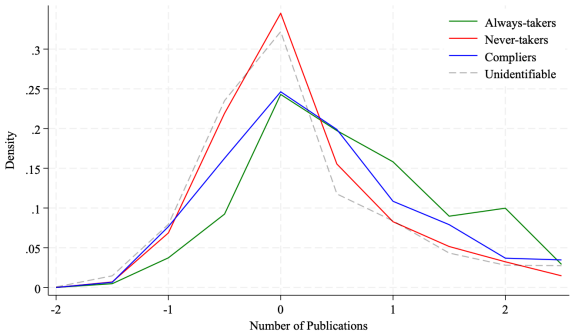
Notes: The figure shows the distribution of instrumental-variable subpopulations – compliers, always-takers, and never-takers – over candidates' publication counts in the year prior to the exam. Publication counts are normalized within exam. The fourth category, *unidentifiable*, denotes candidates in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

Figure E2: Productivity distribution by subpopulation and gender

A) Women

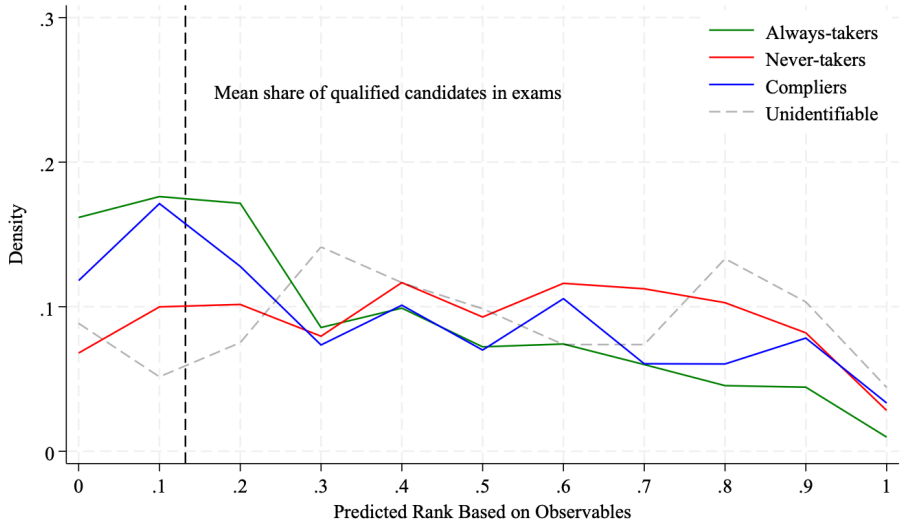


B) Men



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.

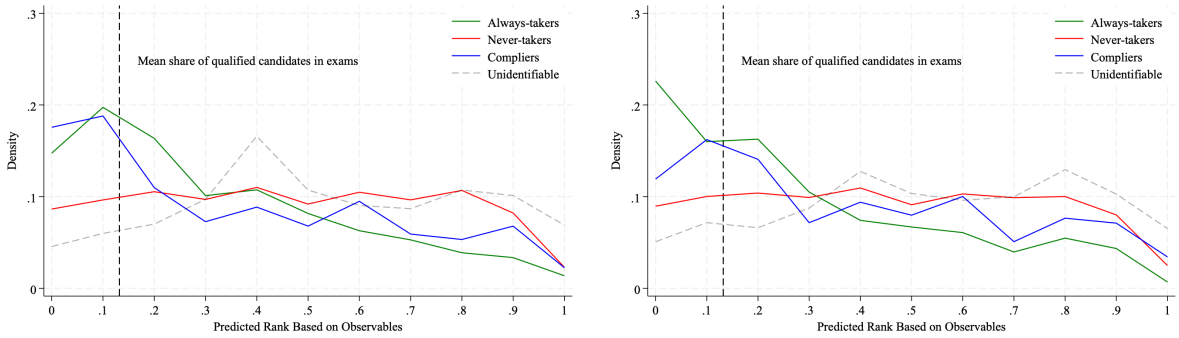
Figure E3: Predicted qualification rank distribution by subpopulation



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 E4: 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: 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 E2: 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 department-level results that complement the main analysis. We first analyze aggregate, non-gender-specific qualification shocks at the Associate Professor and Full Professor levels, paralleling the individual-level evidence in the main text. We then examine how female Associate Professor qualifications affect the stock of tenured women in a department, complementing the main results on promotion flows, and report the corresponding estimates for female Full Professor qualifications. Finally, we further explore the role of research-topic feminization.

F.1 Additional Department-Level Promotion Results

Appendix Figure F1 reports aggregate, non-gender-specific department-level effects of qualification shocks. Consistent with the individual-level results, Associate Professor qualifications generate persistent increases in cumulative promotions to the same rank, whereas Full Professor qualifications generate only a short-run increase that fades over time.

Appendix Figure F2 complements the main promotion-flow results by examining stocks of tenured women. Female Associate Professor qualifications generate a sustained increase in the number of tenured women in the department, although the estimates are less precise than those for promotion flows. By contrast, female Full Professor qualifications have no clear effect on the stock of tenured women.

Appendix Figures F3 and F4 report the corresponding department-level effects of Full Professor qualification shocks. Consistent with the fading individual-level effects at this rank, Full Professor qualifications generate at most short-run effects on Full Professor promotions and no systematic effects on subsequent Associate Professor promotions.

F.2 Research Orientation

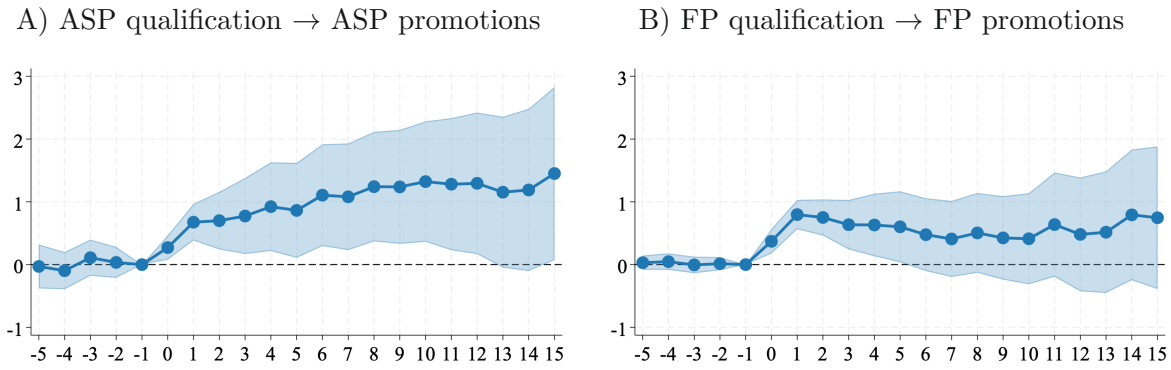
Appendix Table F1 complements the main analysis of qualifications in female-intensive research areas by reporting effects on male outcomes. Across male promotions, male PhD

graduates, male research-active PhD graduates, and male PhD graduates later promoted to Associate Professor, the estimates are small and imprecise.

The table also reports the department-level effects of qualifying candidates working in male-intensive research areas. Unlike qualifications in female-intensive areas, these shocks generate no clear effects on female or male promotions or PhD outcomes.

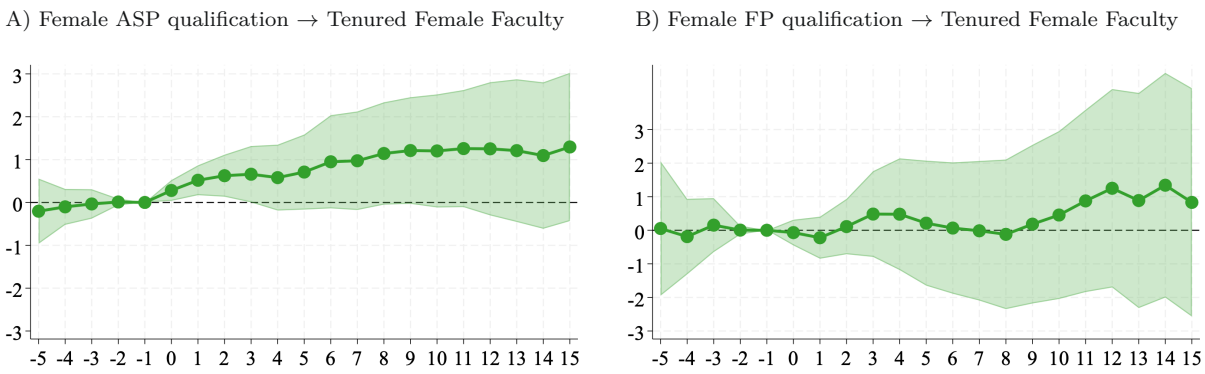
Appendix Table F2 reports horse-race specifications that include both female qualifications and qualifications in female-intensive research areas. For faculty promotions, the coefficient on female-intensive qualifications remains positive, while the coefficient on female qualifications is positive but less precise. For PhD outcomes, the effect of female qualifications on female PhD graduates remains large after controlling for research orientation, whereas the coefficient on female-intensive qualifications is smaller and imprecise.

Figure F1: Impact of exogenous qualifications on the department



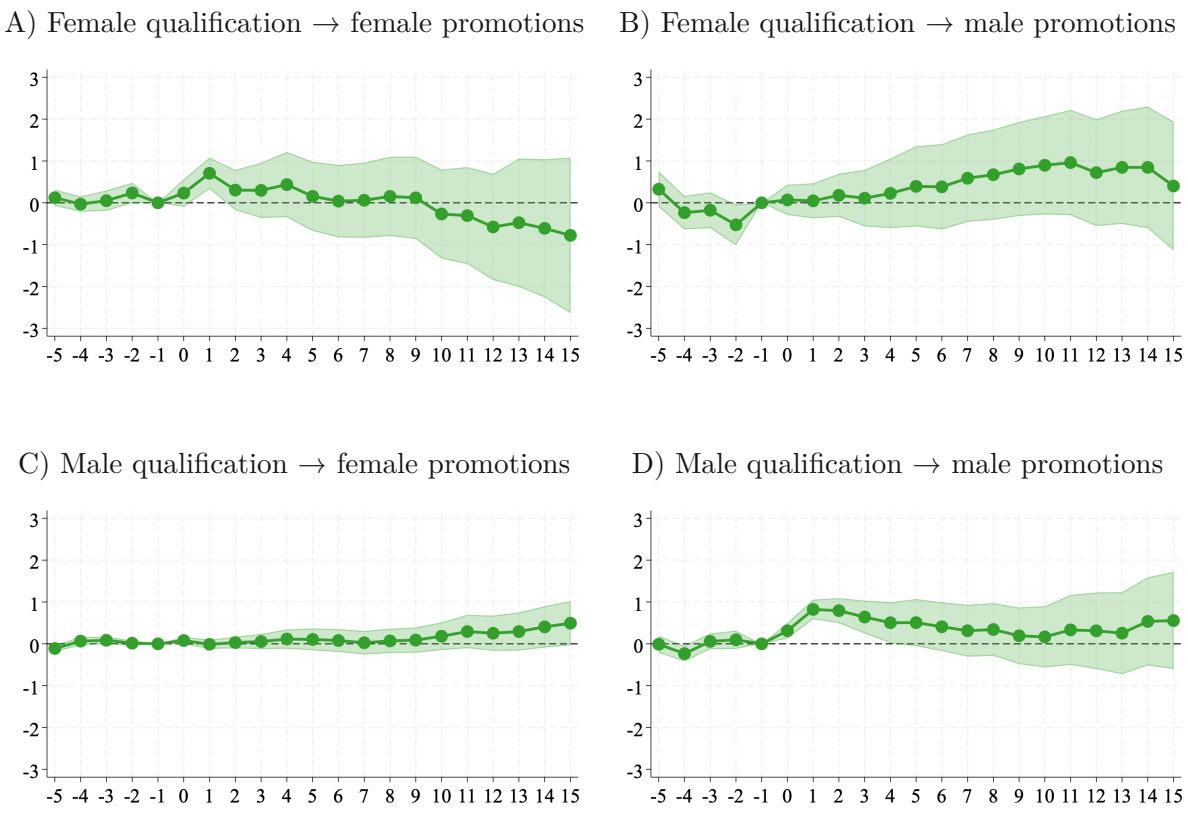
Notes: The figure shows 2SLS estimates of coefficient δ_1 from Equation 4, capturing the impact of exogenously qualifying an Associate Professor on the cumulative number of promotions to Associate Professor (Panel A), and the impact of exogenously qualifying a Full Professor on the cumulative number 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 a female qualification on tenured female faculty



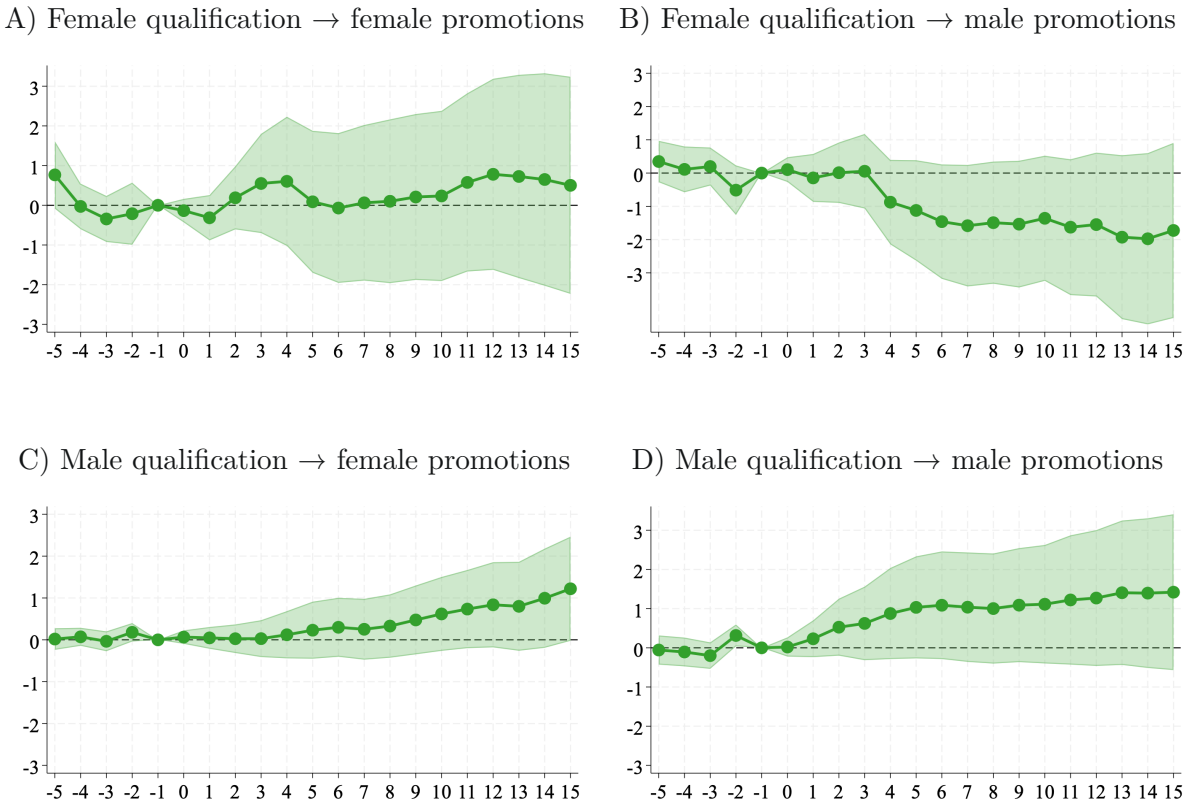
Notes: The figure reports 2SLS estimates of coefficient δ_1 from Equation (6), capturing the impact of exogenously qualifying a woman on departmental faculty outcomes. Panel A shows the effect of qualifying a woman to Associate Professor on the stock of tenured female faculty. Panel B shows the effect 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 4.2. Standard errors are clustered at the department level.

Figure F3: 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 4.2. Standard errors are clustered at the department level.

Figure F4: Department-level effects of female and male qualifications to *Full Professor* on promotions to *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 4.2. Standard errors are clustered at the department level.

Table F1: Department-level effects of qualifications to Associate Professor of researchers in feminized and non-feminized fields

	Department outcomes 10 years following the exam			
	ASP Promotions	Graduates	Research Active Graduates	Graduates Promoted to ASP
	(1)	(2)	(3)	(4)
Panel A. Outcomes of Women				
Feminized researcher qualified	1.044**	3.656*	1.842	0.490**
	(0.411)	(1.967)	(1.126)	(0.247)
Non-feminized researcher qualified	0.027	-2.186	-0.775	-0.052
	(0.545)	(2.787)	(1.638)	(0.352)
Panel B. Outcomes of Men				
Feminized researcher qualified	0.796	1.600	1.204	0.173
	(0.507)	(1.917)	(1.124)	(0.297)
Non-feminized researcher qualified	0.126	-1.021	0.508	0.279
	(0.754)	(3.320)	(1.940)	(0.460)

Notes: The table 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 4.2. The unit of observation is the department–exam cell 6176 observations), and standard errors are clustered at the department level.

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

Table F2: Horse race specification

	Department outcomes of women 10 years following the exam			
	ASP Promotions	Graduates	Research Active Graduates	Graduates Promoted to ASP
	(1)	(2)	(3)	(4)
Panel A. Individual effect of promoting a woman				
Female researcher qualified	1.713**	6.039**	3.029*	0.992**
	(0.671)	(2.902)	(1.695)	(0.396)
Panel B. Individual effect of promoting a researcher in female-intensive area				
Feminized researcher qualified	1.044**	3.656*	1.842	0.490**
	(0.411)	(1.967)	(1.126)	(0.247)
Panel C. Conditional effect of promotion				
Female researcher qualified	1.214	4.558	2.212	0.885**
	(0.763)	(3.369)	(1.943)	(0.444)
Feminized researcher qualified	0.550	1.521	0.849	0.118
	(0.445)	(2.204)	(1.241)	(0.277)

Notes: The table reports 2SLS estimates of coefficients δ_1 and 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 coefficients captures the impact of exogenously qualifying such applicants to *Associate Professor*. Outcomes are measured ten years after qualification and are cumulative over time. The unit of observation is the department-exam cell (6176 observations). All specifications include controls listed in Section 4.2. Standard errors are clustered at the department level.
*** p < .01, ** p < .05, * p < .1

Appendix G — Data Sources and Data Cleaning

We collected data from several sources: the Spanish State Bulletin (*Boletín Oficial del Estado*, BOE) for academic promotions; TESEO, the official registry of PhD dissertations, for doctoral graduates and supervisors; administrative records from the centralized *habilitación* exams; and OpenAlex, Scopus, and Dialnet for research output. This appendix describes how we harmonize these sources. We first summarize the raw data sources, then explain how we construct unique researcher identifiers, impute missing candidate affiliations, link researchers to publication records, deduplicate publications across bibliographic databases, and classify research topics.

G.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*).[36] 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 classification system, which includes 24 two-digit fields, 250 four-digit subfields, and 2,153 six-digit specializations.[37]

National qualification exams. We use administrative data from the centralized national qualification exams conducted between 2002 and 2008 under the *habilitación* system. The records include the universe of candidates and eligible evaluators, the random draws determining the composition of evaluation committees, and the final qualification outcomes.[38] Candidate affiliations are not directly observed in the raw *habilitación* records; below we describe how we reconstruct them.

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

36. 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).

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

38. 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.

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.

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.

G.2 Constructing researcher identifiers

We construct a unique researcher identifier linking individuals across the *habilitación*, BOE, TESEO, and bibliographic data. The procedure is designed to accommodate Spanish naming conventions, frequent surnames and forenames, changes in affiliation over time, and variation in how names are recorded across sources.

When a unique administrative identifier is available, we retain it. For records without an identifier, we use a sequence of exact and fuzzy name-matching routines. The algorithm is augmented with auxiliary information on institutional affiliation, field of specialization, and year of graduation. Before matching, all names are standardized by converting text to uppercase ASCII, removing punctuation and particles, expanding common abbreviations (for example, "Mª" to "MARIA" and "FCO" to FRANCISCO"), correcting common OCR-type errors, and stripping titles.

We first match unidentified records to records with known identifiers. Exact matches on the standardized full name account for the majority of successful links. Remaining records are matched using transformations that reflect common Spanish naming patterns: adding or removing frequent first names, dropping second names, matching on first name plus full

surname, full name plus first surname, initials plus surnames, and related variants. We then match still-unidentified records to one another using the same routines and consolidate transitive chains of matches into common identifiers. Records that were never matched to any other unidentified researcher are retained as singletons.

Several safeguards limit false matches. We use gender as a blocking variable when adding common first names. Matches relying on partial name information are rejected when both surnames are highly common. Potential many-to-one and one-to-many matches are checked using discipline, university, event timing, year of graduation, and available national identifiers. Cases flagged by consistency checks—for example, identifiers mapping to more than two distinct graduation years—are manually inspected and split when necessary. Thesis supervisors and committee members, who are not always part of the main administrative backbone, are then linked using name, event, university, and discipline, with remaining ambiguities resolved manually.

The resulting master researcher file contains, for each researcher-event, the unified identifier, year, academic age, gender, university, academic rank, discipline, and source dataset. We validated the procedure through extensive manual checks of ambiguous and high-risk matches; in these audits, the estimated matching error is below 0.5%.

G.3 Linking researchers to publication records

After constructing the master researcher file, we link researchers to author profiles in Dialnet, Scopus, and OpenAlex. The linking procedure is implemented separately for each database but follows the same general logic: candidate links are generated using exact and fuzzy name matching; candidate pairs are enriched with auxiliary information; matches are accepted through a hierarchy of increasingly permissive rules; and residual many-to-one or one-to-many cases are resolved using discipline, affiliation, timing, and manual inspection.

Dialnet. We link researchers to Dialnet author profiles using exact and fuzzy name matches, combined with Dialnet information on university affiliation, discipline, academic area, and

publication years. Administrative information from our data includes promotion universities, PhD universities, *habilitación* affiliations for eligible evaluators, research fields, and broad disciplines. Early matching steps require agreement on university affiliation, academic area or discipline, an informative name match, and a plausible publication window. Later steps relax the affiliation requirement when Dialnet reports no university, accept broader discipline agreement, and allow one-to-one matches with common names when no available variable conflicts. Many-to-one and one-to-many cases require stronger support, typically agreement on affiliation or discipline combined with uncommon surnames and multiple publications. Remaining conflicts are resolved using discipline, area, and manual inspection.

Scopus. We link researchers to Scopus author profiles using targeted Scopus lookups based on plausible name variants. Candidate researcher–Scopus pairs are enriched with information on administrative affiliation, discipline, research field, promotion or qualification year, and PhD graduation year. On the Scopus side, we use author name strings, publication years, affiliation strings, and paper disciplines inferred from MAG-based codes. Matches are accepted when the publication timeline is plausible and corroborating evidence accumulates across affiliation, discipline, name informativeness, and publication history. Confirmed matches are used to drop competing candidates, enforcing a one-to-one mapping where possible. Residual conflicts are resolved by retaining the link that best fits event timing, maximizes overlap between administrative and Scopus discipline information, and contains richer name and publication evidence.

OpenAlex. We link researchers to OpenAlex author profiles using exact and fuzzy name matching. Match pairs are first screened for ORCID or Scopus identifier conflicts and for name incompatibilities. We then enrich OpenAlex records with Spanish public university affiliations, publication activity years, and journal ISSNs, and compare these to administrative affiliations, PhD and promotion universities, *habilitación* affiliations, discipline, and Dialnet journal information. Matching proceeds in tiers ordered by name-match quality: exact-string matches, substring-based fuzzy matches, and initial-based matches. Within each tier, candidate pairs are confirmed using identifier agreement, affiliation and discipline overlap,

journal ISSN overlap, narrow discipline agreement, and plausible activity windows. Profiles assigned to multiple researchers are resolved in favor of higher-priority tiers or stronger corroborating evidence. When a single researcher is linked to multiple OpenAlex profiles, these profiles are grouped into a composite identifier.

Appendix Table G1 shows the share of researchers matched across databases.

Table G1: 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 given 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.

G.4 Imputing candidate affiliations in the *habilitación* records

Candidate affiliations are not directly observed in the *habilitación* records. We therefore reconstruct them using a sequential procedure that prioritizes administrative information and then uses publication-based evidence.

First, we use links between candidates and other administrative records. Full Professor candidates often also appear as evaluators in Associate Professor exams, which allows us to recover their university at the time they served as evaluators. For early-career candidates who do not yet appear in BOE promotion records, we use the PhD institution when the doctorate was awarded in the exam year or the preceding year.

Second, we use candidates' publication affiliations in Scopus and OpenAlex during the five years before the exam and, when available, the year immediately after the exam call (and

almost certainly before publication of results). We assign the Spanish public university affiliation observed closest to the exam date. When multiple affiliations are observed within this window, we retain the one associated with the largest number of publications, giving priority to affiliations observed closest to the exam. If a publication lists both a public university and a research center, we retain the university affiliation, restricting the set of possible institutions to the 48 Spanish public universities in our harmonized university list.

Finally, if neither administrative sources nor Scopus/OpenAlex publications yield a match, we use Dialnet as a last-resort source. Dialnet reports researchers' most recent institutional affiliation; we retain this information only when it corresponds to a Spanish public university, and otherwise classify the affiliation as private, foreign, or unknown.

G.5 Deduplicating publication records

OpenAlex, Scopus, and Dialnet overlap in coverage. We deduplicate publication records across the three databases to avoid double-counting research output. The procedure identifies duplicate titles within researcher-year windows using string similarity and publication metadata.

We first normalize titles by lowercasing, removing punctuation, diacritics, LaTeX formatting, HTML tags, and stopwords in English and Spanish, and converting Greek characters to textual equivalents. We then apply a multi-step fuzzy-matching algorithm using *token_set_ratio* and *partial_ratio* from the Python *RapidFuzz* library. Candidate duplicates are searched within individual identifiers and publication years, allowing a one-year window to account for discrepancies across databases. Matching proceeds from stricter to more flexible criteria: exact ISSN matches, print/electronic ISSN crosswalks, and matches without ISSN information when title similarity is sufficiently high.

Because pairwise deduplication can generate transitive chains, we identify connected components in the duplicate graph and assign a single canonical publication identifier to each component. Within each duplicate cluster, we retain a representative record, prioritizing OpenAlex, then Scopus, then Dialnet. Unmatched records retain their original identifiers.

Appendix Table G2 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 G1 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 G2: 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 G.6). Column 6 reports the share of publications in Spanish.

G.6 Research topic classification

We classify publications and dissertations into research topics using the OpenAlex topic taxonomy. OpenAlex records contain both *topics* from the OpenAlex taxonomy and *concepts*

inherited from the Microsoft Academic Graph (MAG) (Wang et al., 2019; Hafner and Hedtrich, 2024). Scopus, Dialnet, and TESEO records do not contain OpenAlex topics, so we assign topics to these records using MAG concepts predicted from titles and abstracts.[39] 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 G3 shows the mapping from domains to fields. The complete taxonomy is available at https://api.openalex.org/topics.

Table G3: 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.

To project MAG concepts into OpenAlex topics, we estimate a concept–topic mapping using OpenAlex work–topic–concept triples. We keep concept assignments with concept score ≥ 0.4

39. For Dialnet, we use English-translated titles.

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 $p_{i,t}$.