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

Manuel Bagues Milan Makany Giulia Vattuone Natalia Zinovyeva*

September 22, 2026

Abstract

We study how promotion decisions shape women's careers and the gender composition of academia, exploiting a natural experiment in Spain. Between 2002 and 2008, promotion to tenured positions required passing a national evaluation assessed by randomly drawn committees. Applicants who drew an evaluator with whom they had an academic connection were more likely to pass, generating quasi-random variation in promotions. Using data on about 31,000 applications across 4,000 departments, we document two main findings. First, missing out on tenure has far more persistent consequences for women: failing because of an unfavorable committee draw reduces the probability of holding tenure ten years later by 82 percentage points for women, compared with 39 for men. Second, promoting a woman leads to a lasting increase in women's presence in her department, among both faculty and PhD graduates. Promoting researchers in female-intensive subfields has similar effects, suggesting that these spillovers operate partly through research orientation.

JEL Codes: I23, J16, J44, M51

Keywords: Academic Promotions, Women in Academia, Gender Spillovers, Evaluation Committees

*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 Anna Adamecz, 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 the University of Warwick, 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, KTI-KRTK, University of Bath, University of Vienna, Diversity and Human Capital Workshop at the University of Exeter, the RFBerlin-CEPR Annual Symposium, the EALE Annual Conference, and the SITE Workshop at Stanford University, for useful comments. Natalia Zinovyeva gratefully acknowledges support from the CAGE Research Centre, funded by the ESRC.

1 Introduction

Despite decades of growth in women's doctoral attainment, women remain underrepresented in the upper ranks of academia (Lundberg and Stearns, 2019; Iaria et al., 2022). 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.[1] The lack of women is most acute in science, technology, engineering, and mathematics (STEM), where fewer than one in six full professors are women, but even in the social sciences and humanities the share is below 40%. This issue is particularly salient in economics, where women account for only 23% of full professors in Europe and 18% in the United States, closer to STEM than to the rest of the social sciences (Auriol et al., 2020; Tesar, 2025). Beyond their equity implications, these gaps represent a misallocation of talent with significant consequences for scientific progress and economic growth (Hsieh et al., 2019).

This paper studies the consequences of promotion decisions for women, both for their own careers and for the departments where they work. At the individual level, we ask whether narrowly missing out on tenure has more lasting consequences for women's academic careers than for men's. The tenure decision may arrive at a stage of the career when the opportunity cost of trying again is higher for women, who face a narrower window for childbearing and disproportionately bear caregiving responsibilities (Goldin, 2021; Bütikofer et al., 2018; Kim and Moser, 2025). Women may also be more inclined to interpret an unfavorable outcome as a signal about their chances of success and to give up after a setback, a pattern documented in other competitive settings (Buser and Yuan, 2019; Wasserman, 2023; Kang et al., 2024; Aneja et al., 2026).

At the department level, we ask whether promoting a woman who narrowly qualifies changes the long-run gender composition of her department and of the PhD students it trains, and whether those students remain in academia after graduation. A woman's promotion changes who holds senior positions, which may in turn affect junior women and female students in

1. Source: European Commission (2024), U.S. National Science Foundation (2024), and U.S. Department of Education (2022).

her department through mentoring (Ginther et al., 2020), role models (Carrell et al., 2010; Porter and Serra, 2020), professional networks, or workplace climate (Folke and Rickne, 2022). The potential existence of such trickle-down effects has motivated a range of policies aimed at increasing female representation at senior ranks, from targeted hiring to gender-balance targets.[2]

To answer these questions, we exploit a large-scale 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. This qualification stage was highly selective: the number of qualifications for each field and rank was capped, and only about one in ten applicants qualified. Given the resulting scarcity of qualified candidates, the vast majority of those who qualified were subsequently promoted by their departments. In practice, the *habilitación* system outsourced the promotion decision to the national qualification committees, leaving departments with little more than a formal ratification role. This changed in 2008, when the *habilitación* was replaced by a new system, known as the *acreditación*, in which qualification remained necessary but was no longer capped, restoring universities' discretion over promotions.

The design of the *habilitación* process allows us to identify exogenous variation in qualification outcomes. Candidates seeking promotion 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. Because no conflict-of-interest rules were implemented, many candidates were randomly assigned committees that included a connection: a coauthor, colleague, or PhD advisor. Almost 88% of applicants had at least one connection in the pool of eligible evaluators, but only about one third drew one onto their committee. As Zinovyeva and Bagues (2015) show, this raised their probability of qualifying by roughly 6 percentage points, from a baseline

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

of 11%. Our empirical strategy compares candidates with the same number of connections in the pool of eligible evaluators — and hence the same expected number of connections on the evaluation committee — but whose committees contained more or fewer connections than expected, due to the luck of the draw.

We use information on 17,799 applications to Associate Professor and 13,444 to Full Professor across all academic disciplines. We link these data to a novel dataset covering Spain's public university system: all promotions to senior academic ranks, the universe of PhD dissertations, and publication records from multiple bibliographic sources. This allows us to follow both individual candidates and the roughly 4,000 departments with at least one candidate, tracking faculty composition and PhD student outcomes for up to fifteen years after the initial qualification decisions.

We first examine the individual consequences of qualification. Consistent with the scarcity of qualified candidates, about 90% of those who qualified for Associate or Full Professor were promoted within two years. Many of the candidates who narrowly missed out due to an unfavorable committee draw gradually caught up as they reapplied, but the extent of catch-up differed sharply by rank and, among Associate Professor candidates, by gender. Ten years after the exam, women who narrowly missed Associate Professor qualification were still 82 percentage points less likely to hold a tenured position than women who narrowly qualified, compared with a 39-point gap for men. This asymmetry partly reflects how candidates responded to failure: unlucky women were less likely than unlucky men to reapply, and only among women did research activity decline significantly.

At the Full Professor level we find no comparable asymmetry. Unlucky candidates of both genders largely caught up within ten years. The two ranks differ in two respects that may explain the contrast. Full Professor candidates already held tenured Associate Professor positions, so failure to qualify delayed their career rather than ending it. At the Associate Professor margin, by contrast, unsuccessful candidates had no such position to fall back on, and failure left their continued academic employment uncertain. The application to Associate Professor also comes earlier, when the opportunity cost of trying again is likely to be higher for women.

We then examine whether promoting a woman changes the gender composition of her department, focusing on Associate Professor exams since Full Professor shocks dissipate as unlucky candidates catch up. We aggregate the same lottery variation to the department level. Because departments typically have only one or two candidates per exam, we essentially compare departments in which a female candidate happened to draw a connected evaluator — and was therefore more likely to be promoted — with departments whose female candidates had the same ex-ante chances of drawing a connection but were unlucky. In the short run, a favorable draw simply adds one senior woman to the department: the promoted candidate herself. Over time, two further forces shape the difference between lucky and unlucky departments. As before, unsuccessful candidates gradually catch up, which shrinks it. In addition, the promotion may affect the chances that other women in the department are promoted — raising them, for instance through mentoring or role models, or lowering them if a recent female promotion reduces the scope for further ones. The evidence suggests that the department's response reinforces the initial promotion rather than offsetting it. Ten years after the exam, lucky departments have promoted roughly 1.4 more women, including the promoted candidate herself.

The effects reach further down the pipeline, to the department's doctoral students. Promoting a woman leads to approximately six additional female PhD graduates over a ten-year horizon, three additional female graduates who are research-active five or more years after graduation, and one additional female graduate who reaches a tenured position within ten years of graduation. These effects are large relative to the baseline: the average department graduates about 27 students over the same period, of whom 13 are women. The additional graduates also appear to be at least as strong as other students: about half of them remain research-active and one in six reaches a tenured position, compared with 54% and 4% among female graduates in the average department. We find no comparable effects on male PhD graduates or their subsequent outcomes, nor analogous downstream effects from promoting a man. The expansion in doctoral training therefore appears to be specific to women: it arises when a woman is promoted, and among female students.

Because we observe completed dissertations rather than enrollments, we cannot directly disentangle whether these gains reflect more women starting a PhD or fewer abandoning it along the way. The timing is informative, however: the effects on female graduates who remain research-active and who are later promoted are already significant about three years after the promotion, before most students entering afterwards could have completed their doctorate, which points to completion or retention among students already enrolled. Furthermore, only one third of the additional graduates appear to have been advised by women, which suggests that the effect does not run mainly through direct supervision. Together, these patterns point to broad department-level channels — such as mentoring, role-model exposure, professional networks, and workplace climate — through which senior women make academic careers more attractive and sustainable for junior women. Heterogeneity analysis suggests that some female presence may be a precondition: effects are close to zero in departments that had no senior women before the exam, and largest where women were already present but still a minority. These subgroup differences are not statistically significant, however.

At both the individual and the department level, our analysis identifies the effect of qualification for candidates whose success hinged on drawing a connected evaluator (i.e., compliers). The evidence suggests that their academic records were closer to those of always-takers, who qualified regardless of committee composition, than to those of never-takers, who failed even with a favorable draw. This pattern is to be expected in such a competitive process: a third of candidates drew a connection, but only around one in ten qualified, so a favorable draw could push over the threshold only those already close to it. That threshold was also unusually high by historical standards: during the *habilitación* period, annual promotions to Associate Professor were about a third of their level before and after it, so many compliers would plausibly have been promoted in other periods. Our estimates therefore speak to the effects of reducing promotion frictions for relatively strong candidates.

Finally, we ask whether the department-level effects reflect the gender of the promoted candidate or the research areas in which women tend to work. Because women and men are unevenly distributed across subfields and topics (Dolado et al., 2012; Kozlowski et al., 2022), promoting a woman may also shift the department's research orientation toward areas where women are better represented, which may itself attract female students and faculty. To explore this channel, we estimate the effects of promoting researchers working in *female-intensive* subfields — those where the share of female authors lies above the mean in their field — regardless of the candidate's own gender. Promoting such a candidate generates gains in subsequent female promotions comparable in magnitude to promoting a woman directly. It also increases the number of female PhD graduates by about four over ten years, although this estimate is only marginally significant. Research orientation therefore appears to be an important channel through which promotions shape the presence of women in academic departments. Because it can be implemented without reference to candidates' gender, supporting female-intensive subfields may offer an alternative policy instrument where gender-based measures are constrained.

This paper contributes to several strands of research. First, it adds to the literature on the barriers women face in academic careers. Prior work documents gender differences in credit attribution, publication standards, and professional recognition (Sarsons et al., 2021; Card et al., 2020; Koffi, 2025). Our results suggest that the way the academic career is structured may itself affect women differently: failing to obtain tenure has far more persistent consequences for female candidates. These findings echo evidence from other competitive settings that early setbacks have more lasting consequences for women (Buser and Yuan, 2019; Wasserman, 2023; Kang et al., 2024). The closest parallel is Aneja et al. (2026), who exploit the quasi-random assignment of patent applications to examiners and find that women are less likely than men to continue after an early rejection. The absence of a comparable gap among Full Professor candidates, who already hold a tenured position, suggests that such differences in responses to setbacks matter most when failure is costly, as it is at the tenure stage. A practical implication is that departments and institutions seeking to increase female representation may want to pay particular attention to this career stage, where a single unfavorable outcome can have consequences that last for decades.

Second, we provide the first large-scale causal evidence on whether senior female representation in academic departments generates downstream effects on junior women. Existing evidence points in this direction but is primarily descriptive: economics departments with more female faculty display stronger female representation among students and junior faculty (Boustan and

Langan, 2019), and female PhD students paired with female advisors are more likely to remain in academia (Hofstra et al., 2022). Such correlations may partly reflect departmental culture, resources, or selection. Our design shows that promoting one woman leads to a persistent increase in the number of women promoted and completing a PhD in that department.

Third, we contribute to understanding why senior female representation in academia matters, and what this implies for policy. Prior work has examined several possible channels: role models and mentoring (Carrell et al., 2010; Porter and Serra, 2020; Ginther et al., 2020), advisor–student matching (Hilmer and Hilmer, 2007; Gaule and Piacentini, 2018; Neumark and Gardecki, 1998), workplace climate (Folke and Rickne, 2022), and gender bias in evaluation, although several studies find that female applicants do not necessarily benefit from being evaluated by women (Bagues and Esteve-Volart, 2010; Bagues et al., 2017b; Deschamps, 2024; Card et al., 2020). Consistent with Bello et al. (2025) and Belot et al. (2023), we find that research orientation also plays an important role. Policies that target research areas rather than gender may therefore offer a complementary lever, potentially less vulnerable to the backlash associated with explicitly gender-based affirmative action.[3]

Fourth, our findings speak to the broader literature on gender and organizations, which studies how female leadership affects hiring, promotion, retention, workplace culture, and networks (Kunze and Miller, 2017; Folke and Rickne, 2022; Cullen and Perez-Truglia, 2023). Because we follow departments for over a decade after the exam, we can show that the effects of senior female representation reach beyond an organization's current members to the next generation it trains. Such intergenerational effects are rarely observable in organizational settings, yet they are central to how gender gaps persist.

2 Institutional Context

This section describes several institutional features that are important to our empirical design. We begin with the structure of academic careers in Spain and the role of departments as the

3. Janys (2024) finds that German departments behave as if subject to an implicit quota of one or two women: once that number is reached, further women are unlikely to be hired. Such a ceiling may limit what explicitly gender-based measures can achieve.

units responsible for hiring and promotion. We then turn to the *habilitación* system, which generates the quasi-random variation in promotions used in the analysis. Finally, because part of our analysis concerns PhD graduates, we explain how doctoral training is organized.

2.1 Academic Ranks, Fields, and Departments

Our analysis covers Spain's public universities, which enrolled more than 90% of university students during our period of study.[4] Academic careers follow a nationally defined ladder of ranks. The two senior civil-service ranks are Full Professor (*Catedrático de Universidad*) and Associate Professor (*Profesor Titular de Universidad*), held by 10% and 31% of academic staff in 2004/2005, respectively. We abbreviate Full Professor as FP and Associate Professor as AP. Two further civil-service ranks, *Catedrático* and *Titular de Escuela Universitaria*, accounted for another 16%. These ranks belonged to the *escuelas universitarias*, schools that offered three-year vocational degrees in fields such as nursing, teacher training, and technical engineering, rather than the longer degrees taught in university faculties; they were absorbed into the faculties once those programs disappeared.[5] The remaining positions were non-civil-servant, mostly untenured teaching contracts (33% of academic staff), which in practice served as the typical appointment for researchers who had not yet obtained tenure.[6]

As in the United States, promotion to Associate Professor brings employment security and a substantial role in departmental decision-making, including hiring and promotion. The route

4. In 2004/2005, the midpoint of the *habilitación* period, Spain had 48 public and 22 private universities, and public universities enrolled 91% of all university students (Instituto Nacional de Estadística, 2005). Private universities lay outside the *habilitación* system.

5. *Catedrático de Escuela Universitaria* accounted for 3% of academic staff and *Titular de Escuela Universitaria* for 14%. Both were tenured, but *Titular de Escuela Universitaria* did not require a doctorate and many holders did not have one. Both ranks were phased out toward the end of our period: the three-year programs they served were replaced by four-year degrees under the Bologna reforms, and the reform that replaced the *habilitación* opened a route into *Profesor Titular de Universidad* for holders with a doctorate. Integration was not automatic — it required a specific accreditation, assessed on teaching and management as well as research — but those who obtained it transferred in their existing posts, without competing for a vacancy. About 10% of the candidates in our sample held one of these posts at the time of the exam.

6. Figures refer to 2004/2005, when Spanish public universities employed about 90,000 academic staff. The teaching contracts are *Profesor Asociado* (28%) and *Profesor Asociado de Ciencias de la Salud* (5%); *Profesor Contratado Doctor*, which carries tenure, accounted for 1.6% and the non-tenured *Profesor Ayudante Doctor* for 1.1%. Source: *Estadística de Personal de las Universidades. Series históricas del personal docente e investigador, curso 2004/2005*, Ministerio de Ciencia, Innovación y Universidades, available at https://estadisticas.ciencia.gob.es/jaxiPx/Datos.htm, last accessed 7 July 2026.

to it, however, is different: while in the United States tenure typically follows a standardized probationary appointment as Assistant Professor and operates as an up-or-out decision, in Spain candidates could reach this stage from a range of prior positions — most often temporary postdoctoral and teaching contracts — and those who were not promoted simply kept their existing contract, although it was usually untenured.[7] For many women, the Associate Professor application, typically in the mid-thirties, coincides with the arrival of children. Using a nationally representative survey of doctorate holders, Appendix C shows that only 10% of female doctorate holders have a child at age 30, about half by age 36, and 73% by age 42. Over the same ages, a gender gap in tenure opens up: by age 40, 45% of male doctorate holders hold a tenured university position, against 33% of female ones.

Every faculty member belongs to one of 188 nationally defined fields (*áreas de conocimiento*). Throughout the paper we define a department as a university–field cell, which in most cases coincides with a formal university department. Where it does not — some formal departments span several fields — hiring and promotion are still typically organized at the field level, so the university–field cell remains the decision-making unit. The reverse case, a single field split across several formal departments, arises only in a few large universities, such as 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: about two-thirds of PhD students had completed their undergraduate degree at the same institution, and nearly three-quarters of faculty worked at the university where they earned their PhD (Cruz Castro et al., 2006). In 2002, the Spanish government introduced the *habilitación* system, a centralized promotion procedure aimed at increasing transparency and meritocracy.[8] The reform separated access to

---

7. A small number of research-intensive departments, concentrated in economics and business, operate internal tenure-track systems modeled on the United States, recruiting through the international job market. Their candidates also had to pass the *habilitación*, though this was typically not an obstacle for them.

8. The reform was enacted by the *Ley Orgánica 6/2001, de 21 de diciembre, de Universidades*, available at https://www.boe.es/buscar/doc.php?id=BOE-A-2001-24515; the *habilitación* procedure was regulated by *Real Decreto 774/2002, de 26 de julio*, available at https://www.boe.es/buscar/doc.php?id=BOE-A-2002-15997,

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

The Ministry issued calls at irregular intervals: once in 2002, when the system was introduced, three times a year in 2003 and 2004, and once a year in 2005 and 2006, after which no further calls were issued. In each call, a single committee assessed all applications for a given field and rank. Candidates went through a sequence of evaluation stages, each of which had to be passed to continue. For Full Professor candidates, these consisted of a presentation of their CV and a research seminar; Associate Professor candidates also gave a teaching demonstration. Committee decisions were made by majority vote. Beyond this general structure, committees set their own criteria.

Four features of the *habilitación* process allow us to isolate exogenous variation in promotions. First, candidates could not select into the process based on the identity of the evaluators. They had twenty days to apply, and only after the deadline had passed did the Ministry draw the committee. Second, the seven members of each committee were selected by lottery from the corresponding pool of eligible evaluators (see details in Appendix B). This pool comprised academic staff at public universities and public research institutes who met minimum research productivity thresholds set by the national research evaluation agency. Serving was generally compulsory, and fewer than 2% of those selected were replaced. Third, no conflict-of-interest rules were implemented, and many candidates were randomly assigned committees that included a connection, such as a coauthor, a colleague, or a PhD advisor.[9] Finally, the number of qualifications available was very limited, roughly one for every ten candidates.[10] Qualification was therefore not simply a certification of minimum quality but a binding step in

last accessed 19 August 2026.

9. 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.

10. Formally, the number of qualifications issued in each field-rank call was set by the Ministry on the basis of the positions that universities reported they expected to open.

the promotion process. Universities could not promote candidates who had not qualified, and those who qualified were typically promoted in the department where they already worked.

The *habilitación* drew sustained criticism: rectors objected that universities could no longer appoint their preferred candidates, and candidates bore substantial costs, traveling repeatedly across the country to be evaluated in person (Carreras, 2007). In 2008 it was replaced by the current *acreditación* system, which retains a national qualification stage but without a cap. Annual qualifications across both ranks increased from about 600 to roughly 2,700, as the share of successful applications rose from about one in ten to nearly seven in ten (García-Ayllón Veintimilla and Espín, 2014). Universities thus regained discretion over promotions. In the new system, assessment also rests on candidates' records alone, without the in-person stages described above, and evaluators are no longer selected by lottery, ruling out our design in later periods.

Promotion activity experienced two major slowdowns during our study period: first between 2004 and 2007, as the scarcity of qualifications under 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 (Panel A of Appendix Figure A1).

2.3 Doctoral Training

Doctoral education in Spain is organized within a nationally regulated, program-based framework.[11] Doctoral candidates are admitted to doctoral programs coordinated by Doctoral Schools, rather than being organized solely around bilateral advisor-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 by universities and national quality-assurance guidelines. While junior non-tenured faculty may

11. 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.

act as co-supervisors, primary supervisory responsibility is typically reserved to faculty at the rank of Associate Professor or above.

Among doctorate holders surveyed by the Spanish National Statistics Institute in 2009, 34% had relied primarily on a public-administration grant, 23% on work as teachers or research assistants, 17% on employment outside academia, and 15% on savings, loans, or family support.[12] Available estimates suggest that drop-out rates were substantial, even among funded students. Buela-Casal et al. (2011) find that, among PhD students holding a grant in 2002–2003, only about 40% had graduated within seven years. The effects of female representation among senior faculty on PhD graduates may therefore operate through completion as well as recruitment.

3 Data

We construct a linked dataset combining information on national qualification exams, faculty careers, departmental composition, PhD outcomes, and research output. We use administrative records from the *habilitación* system and link them to information on promotions, publication records, and doctoral graduates. These data allow us to follow the candidates who took part in the *habilitación* exams held between 2002 and 2008, faculty careers and departmental composition from 1985 to 2023, and doctoral graduates from 1990 to 2017. Appendix G describes the data sources and the linking procedure.

In this section, we first describe the *habilitación* records and how we measure candidate–committee connectedness, and show how connections relate to qualification. We then explain how we track individual-level and department-level outcomes of qualification decisions. We conclude by describing our measures of the gender composition of research areas, which allow us to estimate the effect of promoting an applicant depending on the type of research they do.

12. Own calculations using the microdata of the *Encuesta sobre Recursos Humanos en Ciencia y Tecnología 2009*, Instituto Nacional de Estadística (Instituto Nacional de Estadística, 2010).

***Habilitación* records and connections.** Building on Bagues et al. (2017b), we use *habilitación* records covering 967 exams across 174 research fields: 465 for Associate Professor and 502 for Full Professor.[13] We observe 31,243 applications submitted by 16,529 individuals, mainly reflecting repeated attempts across exam waves and, occasionally, simultaneous applications to closely related fields. For each exam we observe both the pool of eligible evaluators — 114 researchers on average — and the seven members drawn by lottery.

The records report candidates' and evaluators' gender, field, rank, and qualification outcome, as well as university affiliation for evaluators and Full Professor applicants. For Associate Professor applicants, we reconstruct affiliation from administrative records and, when these are unavailable, from the affiliations reported in their publications (Appendix G). Women account for 40% of Associate Professor applications and 27% of Full Professor applications. We also observe whether candidates already held a tenured position at the time of the exam. Holding one was a requirement for Full Professor applicants. Among Associate Professor applicants, only 10% hold one, as *Titular de Escuela Universitaria* (Table 1). The share is similar for women and men applying to the same exam. We code a connection between a candidate and an evaluator when the evaluator was a PhD advisor, a coauthor, or a colleague at the same university at the time of the exam — identified from dissertation records, publication records, and university affiliations respectively. Almost 88% of applicants have at least one connection in the pool of eligible evaluators, but only about one third draw one onto their committee: 29% in Associate Professor exams and 35% in Full Professor exams (Table 1). Most connections arise from shared university affiliation, while coauthorship and advisor–advisee links are less common. The number of connections counts each type separately: if an evaluator is both a colleague and a coauthor, they contribute two connections. The measure therefore increases both with the number of connected evaluators and with the number of links to each of them.

We measure candidates' research output at the time of the exam using publication records from OpenAlex, Scopus, and Dialnet. The information includes the number of journal articles

13. 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.

and their Article Influence Score (AIS),[14] book chapters, and books. Applicants to Full Professor positions have roughly twice as many publications as Associate Professor applicants — 33 on average against 17 — though average journal impact is similar across ranks (Table 1). Women tend to have weaker publication records, their success rate is 1 to 2 percentage points lower at both ranks, and they have slightly fewer connections to committee members (Appendix Table A1).

Connections and qualification. Appendix Table A2 compares the pre-exam research output of applicants with and without a connection to their committee, among all applicants (Panel A), among those who qualified (Panel B), and among those who did not (Panel C). Productivity measures are standardized within exam, so that differences are expressed relative to candidates competing for the same qualification. Looking at the overall population of applicants (Panel A), connected candidates are only slightly more productive than unconnected ones: none of the differences exceeds 0.04 standard deviations, the largest being in citations. As expected, successful candidates are substantially more productive than the average applicant, by around a third of a standard deviation. Among them (Panel B), those with a connection have somewhat weaker journal-publication records than those without, by 0.07 to 0.09 standard deviations. Panel A also shows that connected candidates are about twice as likely to qualify: 18% against 9%.[15] If connections never hurt a candidate, this implies that roughly half of the successful connected candidates in Panel B would not have qualified without their connection. Their average research record, 0.31 standard deviations above the average applicant in number of papers, is therefore a mixture of two groups of similar size. Candidates who would have qualified in any case average 0.38, so a simple back-of-the-envelope calculation would place those who owe their qualification to the draw at

14. 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.

15. This comparison assumes that connected and unconnected candidates are otherwise similar. Reassuringly, it matches the first-stage estimates in Section 5.1, which condition on each candidate's expected number of connections: connected candidates have on average 1.5 connections, so the nine-point gap in qualification rates corresponds to about six points per connection, close to the 6.4 and 5.2 percentage points estimated for Associate and Full Professor exams.

about 0.24.[16] Both are far above the unsuccessful candidates who had a connection (Panel C) and therefore failed despite a favorable draw, at −0.05. The candidates whose qualification depended on the committee draw thus seem to be much closer to those who would have qualified anyway than to those who failed.

Individual career outcomes. We follow candidates' subsequent academic careers for up to fifteen years using promotion records from the Spanish State Bulletin (*Boletín Oficial del Estado*), which report all appointments to Associate and Full Professor positions in Spanish public universities. Qualification translates almost mechanically into promotion: 90% of qualified candidates are promoted within two years and 95% within ten, and 84% of them obtain the position in the same department where they were based at the time of the exam (Table 1). Failing is a setback but not an exit: 46% of unsuccessful candidates reapply in a later *habilitación* call and 54% are promoted within ten years, generating gradual catch-up in promotion rates (Appendix Figure A2).[17] We also measure whether candidates remain academically active, defined as publishing at least one journal article, book, or book chapter in a three-year window centered on the horizon considered; for the ten-year horizon, this means publishing between nine and eleven years after the exam. On that measure, 88% of qualified candidates and 81% of unsuccessful ones are still publishing ten years after the exam.

Department-level faculty outcomes. As explained in Section 2, we define a department as a university–field cell, the unit at which hiring and promotion are typically organized. Using the State Bulletin records, we reconstruct annual department-level senior faculty stocks

16. The calculation is as follows. If connections never hurt a candidate, the qualification rate among unconnected applicants, 9%, also measures the share of connected applicants who would have qualified without a connection. Since 18% of connected applicants qualify, these account for about half of the successful connected candidates, and the remaining half owe their qualification to the draw. The average record of the latter, x, then solves $0.31 = 0.49 \times 0.38 + 0.51 \times x$, which gives $x \approx 0.24$. This calculation tentatively treats connections as random. In the empirical analysis we condition on each candidate's expected number of connections to ensure that this is the case.

17. The second share is larger partly because we observe reapplications only within the *habilitación*, whose last call was in 2006, whereas promotions are observed regardless of the system under which candidates later qualified. Candidates who qualified through the *acreditación* that replaced it therefore appear as promoted but not as reapplying.

and promotion flows by rank and gender from 1985 to 2023.[18] This yields 4,730 departments, roughly half of all possible university–field combinations, since universities typically cover only a subset of fields. Our analysis focuses on the 4,000 departments with at least one faculty member participating in a qualification exam, which employed around 90% of all tenured faculty in public universities during the *habilitación* period.

Table 2 describes the department–exam cells in our sample. In the year before the exam, the average department had 8.9 Associate Professors and 2.8 Full Professors. It had 2.2 candidates in the exam, of whom 0.3 qualified. Over the following decade, it promoted 3.6 people to Associate Professor and 2.1 to Full Professor. Department size is highly skewed, which explains the large standard deviations (Appendix Figure A3), but stable over time: across all Spanish departments, the average had five Associate Professors in 2000 and six in 2023, and about two Full Professors throughout (Panels B and C of Appendix Figure A1). Columns 3–6 report the same statistics separately for Associate and Full Professor exams. Differences between the two are small. Within a department, there are typically only one or two candidates of a given gender per exam (one in around two-thirds of cases, two in around a fifth), and qualifications are rare, especially for women: the average cell has 0.11 qualified women and 0.18 qualified men in Associate Professor exams, and 0.05 against 0.17 in Full Professor exams.

Female representation among senior faculty has risen over time. Among Associate Professors, the share of women rose from 32% in 2000 to 43% in 2023; among Full Professors, from 13% to 28% (Appendix Figure A4). It is lower in science, technology, engineering, mathematics, and medicine (STEM+M) than in the social sciences and humanities (SSH): the median STEM+M professor worked in a department where 25% of tenured faculty were women, compared with 30% in SSH (Appendix Figure A5).

18. 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. This overstates tenure at a department for researchers who leave Spanish public universities before age 65 and understates it for those who stay beyond it. 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 that these cases are rare in practice.

PhD graduates and early-career outcomes. We obtain data on the universe of registered PhD dissertations from TESEO, Spain's official dissertation registry. The data cover 203,542 theses defended in Spanish public universities between 1990 and 2017, after excluding approximately 8% with missing information. The records include graduates' gender, university, research field, graduation year, advisor, and dissertation topic. Women account for 47% of PhD graduates and are substantially more likely than men to have a female advisor: 27% compared with 16% (Appendix Table A3). Among the 30% of graduates whose enrollment year we observe, the average time to completion is approximately five and a half years.

We measure three PhD outcomes at the department level, separately for women and men. The first is the cumulative number of graduates after the qualification shock; the average department graduated about 27 PhD students in the decade following an Associate Professor exam, roughly half of them women (Panel C of Table 2). The second is the number who remain academically active, defined as producing at least one journal article, book, or book chapter five or more years after graduation. Around half of graduates remain active, with women slightly less likely than men to do so (52% against 54%). The third is the number promoted to Associate Professor within ten years of graduation, achieved by about 10% of all PhD graduates, with a wider gender gap: 8% of women against 12% of men (Appendix Table A3).

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.[19] 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 an index of 0.60, while *Evolutionary Game Theory and Cooperation* had an index of 0.22.

19. 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.

For PhD graduates, we use fields according to 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 evolution of the Duncan segregation index. At the six-digit level it stood at roughly 0.40 in the 1990s — 40% of female graduates would have had to switch subfields for the share of women to be the same in every subfield — and had fallen to just above 0.30 by 2015; for the broader two- and four-digit classifications it remained much more stable. Segregation has therefore declined, but it remains sizable.

Building on this index, we classify a candidate as working in a *female-intensive* subfield when their index lies above the mean in their field in the year of the exam. This classification overlaps with gender but is not a proxy for it: 37% of candidates in female-intensive subfields are women, compared with 28% in male-intensive ones.

4 Empirical Strategy

This section presents the instrumental-variables (IV) strategy that we use to identify the causal effect of qualifications on individual careers and departmental outcomes. The approach builds on Zinovyeva and Bagues (2015), who show, using the same *habilitación* exams, that candidates randomly assigned to committees that include their academic connections are significantly more likely to qualify. We compare candidates with the same expected number of connections, given the pool of eligible evaluators they faced, but who happened to draw more or fewer of them onto their committee. Recentering realized connections on their expected value in this way, as formalized by Borusyak and Hull (2023), isolates exogenous shocks to evaluator composition, which are orthogonal by design to any systematic differences in candidate networks or departmental structures. We first use this individual-level design to estimate how exogenous qualification affects subsequent promotion prospects, which both provides the causal effect of qualification on individual careers and motivates the research design at the department level. We then aggregate the variation to the department level to study how exogenous promotions of women and men propagate within departments, affecting faculty composition, the number of PhD graduates, and their subsequent careers.

The following subsections describe the identifying assumptions, specification choices, and what the key coefficients identify.

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 in Section 3, since qualification almost directly translated into promotion during the *habilitación* period. The more informative horizon is the long run, where the estimate measures the extent to which the initial advantage persists, given that unsuccessful candidates often 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, where t is the year in which qualification results are announced, typically one to two years after the exam is called. Let $C_{i,e}$ denote the connectedness between candidate i and their committee, defined in Section 3, measured using the composition of the initially drawn committee, before any potential resignations of committee members. This preserves the exogeneity of the instrument, since resignations may be related to evaluators' ties to candidates. Because they affected only around 2% of committee members, our estimates are practically identical to those based on the final committees. Appendix E reports robustness checks based on alternative definitions of connectedness.

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.

The first-stage equation is:

$$Q_{i,e,t} = \alpha_0 + \alpha_1 Z_{i,e} + \alpha_2 \mathbb{E}[C_{i,e}] + \mathbf{\Phi}' \boldsymbol{X}_{i,t} + \mu_e + \lambda_t + \varepsilon_{i,e,t}, \tag{1}$$

where, to increase precision, $\boldsymbol{X}_{i,t}$ includes the candidate's research output at the time of the exam (journal articles, total AIS, book chapters, and books), μ_e are exam fixed effects, and λ_t are fixed effects for the year in which qualification results are released. The coefficient α_1 captures the causal effect of an additional connection to a committee member on the probability of qualification.

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

$$Y_{i,e,t+k} = \beta_0 + \beta_1 \widehat{Q}_{i,e,t} + \beta_2 \mathbb{E}[C_{i,e}] + \mathbf{\Gamma}' \boldsymbol{X}_{i,t} + \mu_e + \lambda_t + \eta_{i,e,t+k}, \tag{2}$$

where $Y_{i,e,t+k}$ is an outcome measured k years after the qualification decision, such as promotion to the rank applied for, and $\widehat{Q}_{i,e,t}$ is the fitted value from the first stage. We estimate this equation by 2SLS, both on the overall sample and separately by gender. Standard errors are clustered at the exam level.

Identifying assumptions and interpretation. The validity of our instrumental variables approach rests on four standard conditions. First, the instrument must predict qualification, which we document in Section 5.1. Second, the variation in committee composition we exploit should be exogenous and unrelated to candidate characteristics, a condition that holds by design given the random assignment of evaluators and is confirmed empirically in Appendix Table D1. Third, the exclusion restriction requires 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 is unlikely: national evaluations were one-off events with no formal role in university-level hiring, and the promotion process was largely mechanical, since universities faced a tightly constrained number of positions and promoted about 90% of successful candidates within two years of qualification. Fourth, monotonicity requires that connections never reduce a candidate's probability of qualifying, which could happen if a strained professional relationship led an evaluator to be harsher. 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 magnitude of our estimates.

Under these conditions, the coefficient β_1 in equation (2) identifies the Local Average Treatment Effect (LATE) of qualification on promotion for compliers, i.e., candidates who qualify only when they draw a favorable committee. The other candidates are always-takers, who qualify regardless of the draw, and never-takers, who fail even with a favorable one. As shown in Section 3, compliers' research records sit close to those of always-takers and far above those of never-takers. Our estimates therefore capture the effect of qualification for candidates close to the margin, rather than for the average applicant. This estimand is policy-relevant: it concerns candidates strong enough to be competitive but whose outcomes depend on the composition of their committee, who are likely to be the most responsive to interventions that reduce promotion frictions.

Two further considerations bear on the interpretation of the LATE. First, with a continuous instrument, 2SLS recovers a weighted average of effects across margins of the instrument, and these weights need not correspond to a parameter of direct policy interest (Mogstad and Torgovitsky, 2024). We address this concern by showing that our results are robust to alternative instrument definitions, including the binary specification reported in Appendix E.

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 clarifies what the LATE captures: candidates whose qualification is both marginal and promotion-relevant. In practice the scope for such behavior was limited, since the centrally determined cap made qualification itself the primary bottleneck, giving connected evaluators strong incentives to support their candidates regardless of immediate promotion capacity.

4.2 Department-level identification

The department-level analysis builds directly on the individual design, aggregating the same random variation in committee composition to ask how qualification outcomes propagate beyond the individual candidate. 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, regardless of the qualified candidate's gender. 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. We construct the department-level instrument by averaging, across a department's candidates, the deviation between individuals' realized and 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 define the department's average expected connectedness analogously:

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

We average rather than sum individual deviations because a favorable draw can rarely secure more than one qualification for a department. Most connections arise from shared affiliation, and an evaluator from a department's own university is connected to all of its candidates in the exam. Summing deviations would count such an evaluator once per candidate, as if they could help each of them qualify. In practice they rarely help more than one: in 85% of department–exam cells with any qualification, exactly one candidate qualified. The average instead measures how favorable the draw was for the department's candidates as a group.

Letting $Q_{j,e,t}$ denote the number of candidates from department j who qualify in exam e at time t, the first-stage equation is:

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

and the corresponding second stage is:

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

where $Y_{j,e,t+k}$ denotes a departmental outcome measured over the period $[t, t+k]$, such as the cumulative number of promotions or PhD graduates, and $Q_{j,e,t}$ is instrumented by $\overline{Z}_{j,e}$, so that δ_1 captures the causal effect of one additional exogenous qualification on departmental outcomes. Both stages control for the department's average expected connectedness, $\mathbb{E}[\overline{C}_{j,e}]$, so that the comparison is between departments whose candidates faced similar ex-ante chances of drawing connections. The vector $W_{j,t}$ adds the number of applicants, pre-exam department faculty size, pre-exam PhD cohort size, and promotion rates in the years preceding the exam. These additional controls are not required for identification, since the instrument is defined as a deviation from expected connectedness, but 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 conditions similar to those of the individual-level analysis. Exogeneity follows directly from the random assignment of evaluators: the average deviation in committee composition $\overline{Z}_{j,e}$ is by construction unrelated to predetermined departmental characteristics, as confirmed empirically in Appendix Table D2. The exclusion restriction and monotonicity carry over from the individual-level analysis, since a department's qualifications are the sum of its candidates' outcomes.

Impact of qualification shocks, by gender. To separate 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 separate instruments for female and male candidates, averaging the deviation between realized and expected connectedness within each group:

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

with $\overline{Z}^g_{j,e} = 0$ if $N^g_{j,e} = 0$, and gender-specific controls $\mathbb{E}[\overline{C}^g_{j,e}]$ defined analogously and included in both stages. Appendix Table D3 reports balance evidence for these gender-specific instruments.

The first stage instruments the numbers of qualified women and men using both instruments, allowing a department's female connections to affect male qualifications and vice versa:

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

The second stage, for a departmental outcome such as the gender composition of faculty or the career outcomes of PhD graduates, is:

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

The coefficient δ^f_1 captures the effect on departmental outcomes of one additional female qualification, holding constant the number of male qualifications, and δ_1^m the effect of one additional male qualification, holding constant the number of female qualifications. As in the individual-level analysis, both are Local Average Treatment Effects, identified from departments whose number of qualified women or men is shifted by the random composition of the committee.

5 Results

We present the results at the two levels introduced above. At the individual level, Section 5.1 examines how qualification decisions affect candidates' own careers. At the department level, Section 5.2 analyzes how qualification shocks change the subsequent gender composition of departmental faculty, and Section 5.3 studies downstream effects on PhD graduates and their academic outcomes. Two further subsections extend the analysis: Section 5.4 asks whether the effects differ by a department's baseline gender composition and field, and Section 5.5 asks whether they reflect the gender of the promoted candidate or the research areas in which women tend to work. Throughout, we abbreviate percentage points as *p.p.* and standard errors as *s.e.*

5.1 The Effects of Qualification on Candidates' Careers

Academic connections strongly increase qualification probabilities, with effects similar to those in Zinovyeva and Bagues (2015). In Associate Professor exams, each additional connection among the seven evaluators raises the probability of qualification by 6.4 p.p. (s.e. = 0.6), representing a 54% increase relative to the baseline qualification rate (first-stage estimates of equation (1) in Appendix Table D4). In Full Professor exams, the effect is 5.2 p.p. (s.e. = 0.5), or 48%. These estimates are highly significant and virtually unchanged when controlling for prior research productivity (columns 1–2 and 5–6), as expected under random committee assignment. The productivity controls are themselves positively associated with qualification, as one would expect: candidates with more journal articles, a higher total AIS, and more book chapters are more likely to qualify. Importantly for the gender comparisons that follow, women and men benefit similarly from having a connection on their committee: each additional connection raises the probability of qualification by 6.1 p.p. for women and 6.8 p.p. for men in Associate Professor exams, and by 4.0 and 5.5 p.p. in Full Professor exams, with no statistically significant difference between them (columns 3–4 and 7–8). First-stage F-statistics range from 19 to 132 across specifications, indicating that weak instruments are not a concern.

Using this exogenous variation, we examine how qualification affects candidates' long-run careers. Panels A and B of Figure 1 plot 2SLS estimates of β_1 from equation (2), the effect on promotion over a 15-year horizon, separately by rank. These estimates net two forces: qualification translated almost mechanically into promotion, with about 90% of qualified candidates promoted within two years, while unsuccessful candidates could reapply in later calls and gradually catch up. In the short run, effects are similar across ranks. In both cases, one year after the committee decision, exogenously qualified candidates are roughly 70 to 75 p.p. more likely to have been promoted — already short of the near-mechanical 90%, indicating that catch-up begins quickly.

In the long run, the career advantage persists at the Associate Professor level: 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. This incomplete catch-up is considerably more pronounced among women (Panels C and E of Figure 1). Ten years after the qualification decision, exogenously qualified women are 82 p.p. (s.e. = 17) more likely to have been promoted, compared with 39 p.p. (s.e. = 12) for men; the gender difference is marginally significant (p-value = 0.052). Fifteen years after the qualification exam, the promotion gap between candidates who narrowly qualified and those who narrowly missed is still 56 p.p. (s.e. = 15) for women and 30 p.p. (s.e. = 12) for men, though the difference between them is no longer statistically significant (p-value = 0.16).

Catch-up at the Associate Professor level requires that unsuccessful candidates reapply and eventually qualify, so the gender gap could arise at either margin. Table 3 reports the effect of narrowly failing on reapplication to the *habilitación* and on subsequent research activity. Candidates who qualify do not reapply, so the estimated effect of failing equals the reapplication rate of candidates who failed because of an unfavorable committee draw: 42% of these men reapply in a later *habilitación* call, compared with 14% of women, a difference significant at the 5% level. Reapplication is observed only within the *habilitación*, so these rates capture a short window. Research activity shows the same asymmetry. For women, narrowly failing reduces the probability of remaining research-active by 28 p.p. after five years and by 37 p.p. after fifteen, and the effect is significant at every horizon. For men, the estimated reduction is smaller and not significant at the 5% level at any horizon. These patterns suggest that women who narrowly fail are less likely to reapply in the short run and less likely to remain active in research, helping explain the more persistent promotion effects for women. As discussed in the introduction, at least two mechanisms may contribute to this asymmetry. First, women may be less likely than men to try again after a setback, a pattern documented in other competitive settings. In our setting, an unfavorable result may be read as a signal about candidates' future chances of success, and more readily so by women, who, according to survey evidence, tend to assess their own research productivity below that of comparable men (Bagues et al., 2017a).[20] Second, the opportunity cost of trying again is likely to be higher for women at this stage: Associate Professor candidates typically apply in their mid-thirties, when the window for childbearing is narrowing.

At the Full Professor level, the effects on promotion instead decline steadily, become statistically insignificant after around six years, and are close to zero by the end of the 15-year horizon. The stronger catch-up among unsuccessful Full Professor candidates likely reflects their tenured status, which allowed them to reapply without their careers at stake. Moreover, they typically apply in their mid-forties or later, at an age when family-formation constraints are less binding. The same pattern holds for women and men separately (Panels D and F of Figure 1), though the female estimates are imprecise given the small number of female Full Professor candidates.

20. In a survey of about 1,900 Spanish university faculty conducted in 2016, women's self-assessed research productivity was 0.14 standard deviations lower than that of men with similar observable characteristics, including research productivity assessments (*sexenios*), rank, field, and experience.

5.2 The Long-Run Effects of Female Promotions on Faculty Composition

Table 4 reports estimates of the department-level first stage, equation (3), which captures how random variation in committee connections affects how many of a department's candidates qualify; Panel A covers Associate Professor exams and Panel B Full Professor exams. When candidates from a department draw, on average, one additional connection relative to expectations, the number of qualified candidates rises by 0.15 ($s.e. = 0.01$) at the Associate Professor level and 0.11 ($s.e. = 0.01$) at the Full Professor level (column 1).

Three forces shape the second-stage estimates of equation (4): the direct promotion of the qualified candidate, the gradual catch-up of candidates who drew an unfavorable committee and reapplied in later calls, and the department's own longer-run response, which may reinforce the initial promotion or offset it if a recent promotion reduces the scope for further ones. We find that each additional qualification translates into roughly 0.7 additional Associate Professor promotions and 0.8 additional Full Professor promotions in the short run, matching the individual-level estimates (Panels A and B of Appendix Figure F1). Over longer horizons the Associate Professor effect continues to grow, while the Full Professor estimate becomes too imprecise to be informative. The gender-specific analysis that follows therefore focuses on Associate Professor qualifications: Full Professor qualifications left no lasting effect on individual careers, and female Full Professor candidates are too few for a well-powered analysis by gender. The corresponding Full Professor estimates, reported in Appendix Figures F2 and F3, show no significant long-run effects, although they are too imprecise to rule out relatively large effects.

We now split qualifications by the gender of the qualifying candidate, estimating equations (5) and (6). The first stage behaves similarly for both groups: departments whose female or male candidates draw more favorable committees experience comparable increases in the number of qualifications (columns 2 and 3 of Table 4). The instrument remains strong, with a first-stage F-statistic of 34 at the Associate Professor level, against 117 in the pooled specification. Figure 2, which plots δ_1^f and δ_1^m from equation (6), decomposes Associate Professor qualification shocks by candidate gender. 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. The effect then rises over time, reaching approximately 1.4 additional female promotions after ten years ($s.e. = 0.6$), an estimate that is statistically significant at the 5% level, and remaining at a similar level through year fifteen, when it becomes less precise. This is a large effect: over the same ten years, the average department promoted 1.5 women to Associate Professor (Table 2). Because this outcome counts all female promotions in the department, the estimate combines the promotion of the qualifying candidate with any later promotions of other women. It is larger than the candidate's own gain in the probability of promotion at the same horizon — 0.82 at the individual level after ten years — which would be consistent with the department's response reinforcing the initial promotion rather than offsetting it. The difference is imprecisely estimated, however, and we cannot reject that the effect is confined to the qualifying candidate's own promotion.

These gains do not appear to come at the expense of men: female qualification shocks have no significant effect on male Associate Professor promotions at any horizon (Panel B of Figure 2), and male qualification shocks generate no significant effects on either female or male promotions (Panels C and D). The estimates are too imprecise to reject that these effects are all similar, however; the effect of a female qualification on female promotions is the only one distinguishable from zero.[21]

5.3 The Downstream Effects of Female Promotions on PhD Graduates

The previous results show that the qualification of one additional woman by chance produces a lasting increase in female promotions in her department. Figure 3 traces the effects of this initial female qualification on PhD graduates, which capture the combined effect of the promotion of the qualified applicant itself and any later changes in faculty composition that it induced. Departments experiencing an exogenous female qualification train substantially more female PhD graduates over time (top row, left panel). The effect begins to emerge three to four years after qualification and becomes statistically significant around year five. After

21. Appendix Figure F4 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.

ten years, departments have graduated about six additional female PhD students. There is no significant effect on male PhD graduates (second row, left panel), suggesting that the increase does not come from within-department substitution away from male students.

Female PhD graduates in affected departments are also more likely to remain in academia and to be promoted. Following an exogenous female qualification, approximately three additional female graduates remain academically active (top row, center panel), and one additional female graduate is promoted to Associate Professor within ten years of graduation (top row, right panel). Set against the six additional graduates, these imprecise magnitudes suggest that the additional women are at least as strong as other students: about half of them remain research-active and one in six reaches Associate Professor, compared with 54% and 4% among female graduates in the average department (Panel C of Table 2). Female qualification shocks have no significant effect on male graduates' academic activity or later promotion (second row, center and right panels), though the estimates are not precise enough to rule out effects similar to those for women. Male Associate Professor qualification shocks do not generate comparable downstream gains. If anything, they are associated with a modest decline in female PhD outcomes (third row, center and right panels), although these estimates are only marginally significant.

The timing of these effects is informative about the underlying channel. Qualified candidates were typically promoted within a year of the exam, and the median doctorate in our data takes five years, so students entering the department after the promotion would not begin to graduate before year four and would mostly graduate from year six onward. Yet close to half of the ten-year effect is already in place by year five. Part of the increase is therefore likely to operate through students who were already enrolled — through retention, completion, or faster progression — rather than through new recruitment alone. There is ample scope for such effects, given the high attrition in Spanish PhD programs, where only about 40% of grant holders had graduated within seven years. The early estimates for the total number of graduates are individually imprecise, but the timing is sharper for graduates who stay in academia: the effects on female graduates who remain research-active and who are later promoted are already statistically significant about three years after qualification (top row, center and right panels of Figure 3), before most newly recruited students could have completed their doctorate.

Two channels could generate these effects. The effect of a female promotion may operate directly through supervision: primary supervisory responsibility is typically reserved to faculty at the rank of Associate Professor or above (see Section 2), so qualification is what allows the newly promoted woman to take on doctoral students of her own. Alternatively, it may operate through broader departmental changes — mentoring, role models, professional networks, or research orientation — that reach female students regardless of who supervises them. We separate the two by estimating the effects on the number of female graduates supervised by women and by men. Appendix Figure F5 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. Both estimates are significant only at the 10% level at that horizon, so the comparison is suggestive rather than conclusive. The estimates nonetheless suggest that the promoted woman's own supervision cannot explain the full increase, since most of the additional female graduates are supervised by men. This points to broader departmental changes rather than direct supervision: senior faculty sit on the committees that admit doctoral students, assign supervisors, and monitor their progress, and many students depend on departmental contracts for funding, so a female promotion can affect students' completion without running through her own supervision. We examine the research-orientation channel directly in Section 5.5.

Overall, the evidence suggests that promoting a marginal female applicant 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.

5.4 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 is lower on average in science, technology, engineering, mathematics, and medicine (STEM+M) than in the social sciences and humanities (SSH), 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, these estimates have limited statistical power: tests of equality across subgroups may fail to detect differences of moderate size, so the patterns should be interpreted with caution.

Panel A of Appendix Table F1 splits departments by the pre-exam share of women among Associate and Full Professors: no senior women (30% of department-exam observations), women present but in the minority (56%), and women in the majority (14%). The largest effects appear in departments where women are present but remain a minority. In this group, a female qualification generates approximately two additional female Associate Professor promotions and eleven additional female PhD graduates within ten years — effects that are large and statistically significant. By contrast, estimates for departments with no senior women at baseline are imprecise and not statistically distinguishable from zero, and positive but smaller and imprecisely estimated in departments where women are already in the majority. 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 Appendix Table F1 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.

5.5 Gender and Research Orientation

Because research areas differ substantially in their gender composition, promoting someone in a female-intensive subfield could strengthen a department's female pipeline even when the candidate is a man; we therefore examine whether such qualifications generate downstream effects comparable to those of female qualifications. Using the classification of *female-intensive* subfields described in Section 3, we re-estimate equations (5) and (6), replacing the split by candidate gender with a split by research area. The endogenous variables become the number of a department's candidates who qualify in female-intensive and in male-intensive subfields, instrumented by the average deviation of realized from expected connectedness computed separately within each group, with expected connectedness in each group entering as a control in the same way. The coefficient of interest is the analogue of δ_1^f: the effect of one additional qualification in a female-intensive subfield, holding qualifications in male-intensive subfields fixed. Columns 4 and 5 of Table 4 report the corresponding first stage. An additional connection drawn by a department's candidates working in female-intensive subfields raises the number of qualifications in those areas by 0.13 ($s.e. = 0.02$, Panel A), and the analogous effect for male-intensive subfields is 0.11 ($s.e. = 0.02$). The first stage remains strong in this specification, with an F-statistic of 19.

One year after the exam, qualifications in female-intensive subfields generate 0.4 additional female promotions, against 0.6 for female qualifications (Figure 4). The gap then closes: after ten years the effect reaches 1.0 ($s.e. = 0.41$), against 1.4 for female qualifications, and after fifteen 1.6 ($s.e. = 0.58$), against 1.4. Male promotions are largely unaffected, and qualifying candidates in male-intensive subfields has no clear effect on promotions or PhD outcomes of either gender (Appendix Table F2). The one-year effect is close to what one would expect mechanically, since 37% of candidates in these subfields are women; its later growth, despite the catch-up of unsuccessful candidates, suggests that these qualifications also raise the promotion of other women in the department. Research orientation therefore appears to be an important channel for shifting faculty gender composition.

The effects on PhD outcomes are positive but smaller and less precise. Over ten years, qualifying a candidate in a female-intensive subfield generates 3.7 additional female PhD graduates ($s.e. = 2.0$), against six following a female qualification; 1.8 additional graduates who remain research-active ($s.e. = 1.1$), against three; and 0.49 who reach Associate Professor within ten years of graduation ($s.e. = 0.25$), against one. The estimates are individually imprecise and not statistically distinguishable from those for female qualifications (Panels B–D of Figure 4). The effects on male PhD outcomes are limited throughout (Appendix Table F2).

Taken together, the results suggest that research orientation is an important channel through which promotions shape the presence of women in departments, both among faculty and among doctoral students.

6 Conclusions

This paper examines two hypotheses about the underrepresentation of women in senior academic positions: that the structure of academic careers makes setbacks more costly for women, and that the scarcity of senior women is self-perpetuating. We study both hypotheses 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 — through candidates' own careers and through the departments where they work.

Our first result is that failing to qualify for Associate Professor has large and persistent career consequences, especially for women. Narrowly missing qualification reduces women's chances of promotion much more than men's, even ten years later. Part of this asymmetry arises because these women are less likely to reapply and to remain active in research. The pattern is consistent with two explanations: women may be more inclined to read an unfavorable outcome as a signal about their chances of success, and the tenure decision arrives in the years when family decisions are hardest to postpone. In line with this, no comparable gender gap appears among Full Professor candidates, who already hold a tenured position and typically apply in their mid-forties or later.

Our second result is that promoting a woman affects her department beyond her own career: female promotions in the department rise, and the effects extend to the PhD pipeline. Departments experiencing an exogenous female qualification train more female PhD graduates, and these additional graduates are no less likely than others to remain active in research or to reach Associate Professor. The additional graduates are not concentrated among students supervised by women, pointing to broader department-level changes rather than only direct advising relationships. Taken together, the PhD results support the argument that female representation at senior ranks is partly self-reinforcing: adding one woman affects the next generation of women the department trains, not only her own career.

The department-level estimates raise two displacement concerns. Within departments, gains for women could come at the expense of men. We find little evidence of this: female qualification shocks do not significantly reduce male promotions, male PhD graduates, or male graduates' subsequent academic outcomes. Across departments, gains where a woman qualified could be offset by losses elsewhere. Our design does not directly address such general-equilibrium effects, but we do not expect this to be a major concern here. A pure reallocation story would require substantial sorting of female researchers or students across departments in response to local promotion shocks, which is unlikely in a setting with limited geographical mobility. Moreover, part of the effect on PhD graduates appears among students who were already enrolled at the time of the promotion, which points to higher retention, plausible where fewer than half of funded PhD students complete their degree, rather than to students choosing a different department.

A natural question is whether these effects reflect the gender of the promoted candidate or the research they do. Because women and men specialize in different areas, promoting someone in a female-intensive subfield may attract and retain women regardless of the candidate's own gender. Consistent with this hypothesis, we find that qualifying a candidate in a female-intensive subfield eventually raises female faculty promotions by about as much as qualifying a woman, and it also increases the number of female PhD graduates, although this effect is smaller and less precisely estimated.

Because the *habilitación* was highly selective, the candidates whose outcomes our design identifies were accomplished researchers who happened to fall on either side of a chance draw. Our estimates therefore speak to policies that reduce frictions facing strong female candidates. Measures that keep capable women from being derailed by an arbitrary setback — limiting the role of chance in promotion decisions, expanding opportunities to reapply, or providing support after an unsuccessful evaluation — matter on two counts. They protect the careers of the women directly affected, who in our setting are the ones least likely to try again. And because a woman's promotion improves the outcomes of female doctoral students in her department, their effects reach well beyond her own career. Where policies that explicitly condition on candidates' gender are not feasible, supporting female-intensive subfields may offer an alternative policy instrument.

References

Aneja, **Abhay**, **Oren Reshef and Gauri Subramani**. 2026. "Attrition and the Gender Patenting Gap". *The Review of Economics and Statistics* 108 (4), pp. 1017–1026.

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.

Bagues, **Manuel**, **Irma Clots-Figueras and Natalia Zinovyeva**. 2017a. "El uso del tiempo en el mundo académico: diferencias de género". *Newsletter of COSME (Comité Sobre la Situación de la Mujer en Economía)* (5).

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

Bagues, **Manuel F. and Berta Esteve-Volart**. 2010. "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment". *The Review of Economic Studies* 77 (4), pp. 1301–1328.

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).

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, María 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". *Cultura y Educación* 23 (2), pp. 273–283.

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.

Bütikofer, Aline, Sissel Jensen and Kjell G. Salvanes. 2018. "The Role of Parenthood on the Gender Gap among Top Earners". *European Economic Review* 109, pp. 103–123.

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.

Carreras, Josep. 2007. "Evaluación de la calidad docente y promoción del profesorado (V): legislación universitaria española (c): desarrollo de la Ley Orgánica de Universidades (2002-2005)". *Educación Médica* 10 (2), pp. 67–68.

Cruz Castro, Laura, Luis Sanz Menéndez and Jaime Aja Valle. 2006. *Las trayectorias profesionales y académicas de los profesores de universidad y los investigadores del CSIC*. Working Paper. Consejo Superior de Investigaciones Científicas (España).

Cullen, Zoë and Ricardo Perez-Truglia. 2023. "The Old Boys' Club: Schmoozing and the Gender Gap". *American Economic Review* 113 (7), pp. 1703–1740.

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.

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 Llorenç Arguimbau-Vivo. 2010. "Las tesis doctorales en España (1997-2008): análisis, estadísticas y repositorios cooperativos". *Revista Española de Documentación Científica* 33 (1), pp. 63–89.

García-Ayllón Veintimilla, **Salvador and Antonio Tomás Espín**. 2014. “La acreditación y promoción del profesorado en la universidad española: situación, tendencias y perspectivas de futuro”. *REDU. Revista de Docencia Universitaria* 12 (4), pp. 39–62.

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. 205–209.

Goldin, **Claudia**. 2021. *Career and Family: Women’s Century-Long Journey toward Equity*. Princeton University Press.

Hafner, **Flavio and Christoph Hedtrich**. 2024. “Networks in the Market for Researchers”. 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.

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

Instituto Nacional de Estadística. 2005. *Nuevas carreras, más universidades*. Cifras INE. Boletín informativo del INE 4/2005.

— 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.

Kim, **Scott Daewon and Petra Moser**. 2025. “Women in Science: Lessons from the Baby Boom”. *Econometrica* 93 (5), pp. 1521–1560.

Koffi, **Marlène**. 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. "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. 226–254.

Priem, **Jason**, **Heather Piwowar and Richard Orr**. 2022. *OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts*. arXiv:2205.01833.

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).

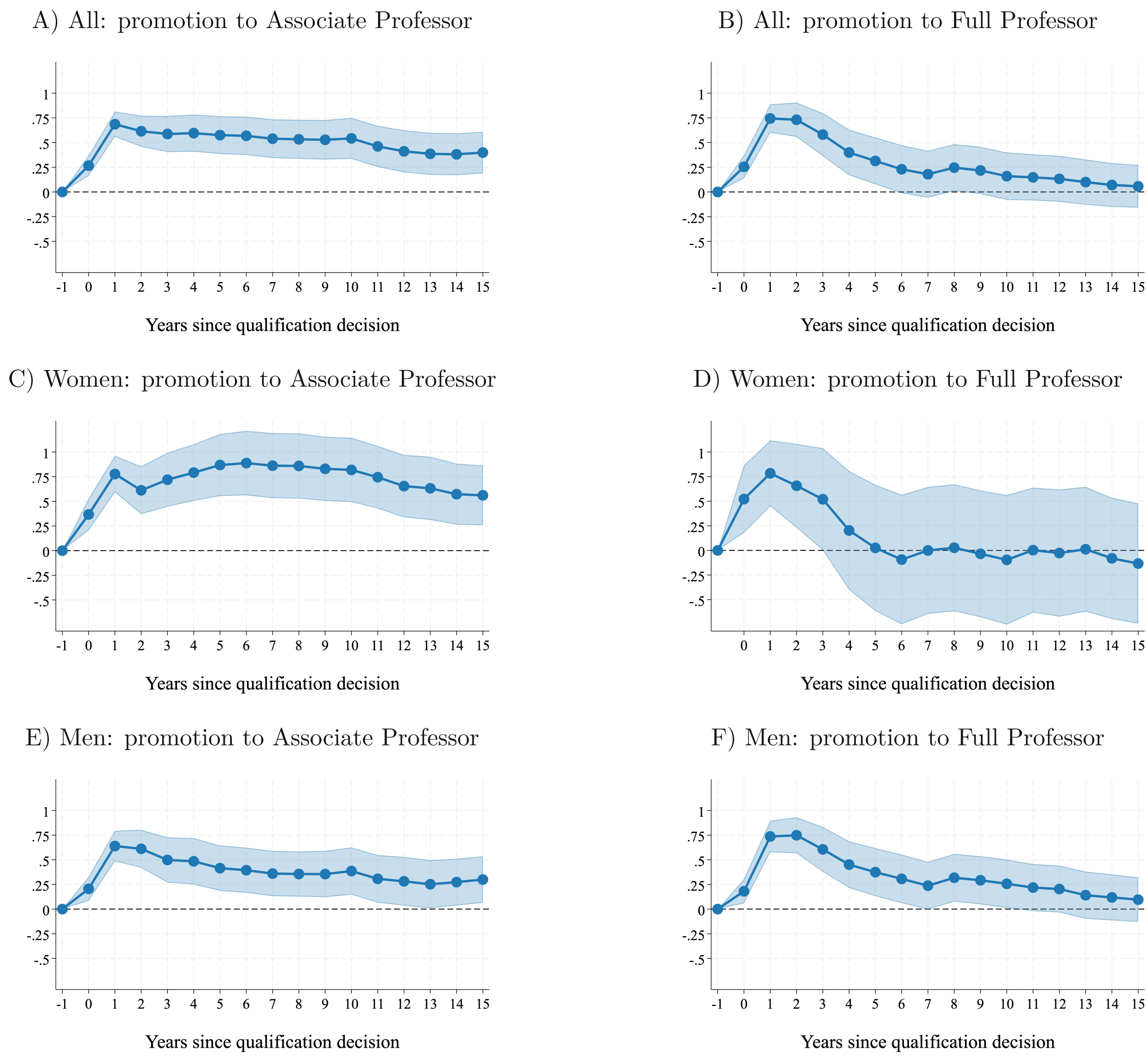
Wang, **Kuansan**, **Zhihong Shen**, **Chiyuan Huang**, **Chieh-Han Wu**, **Darrin Eide**, **Yuxiao Dong**, **Junjie Qian**, **Anshul Kanakia**, **Alvin Chen and Richard Rogahn**. 2019. "A Review of Microsoft Academic Services for Science of Science Studies". *Frontiers in Big Data* 2.

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

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

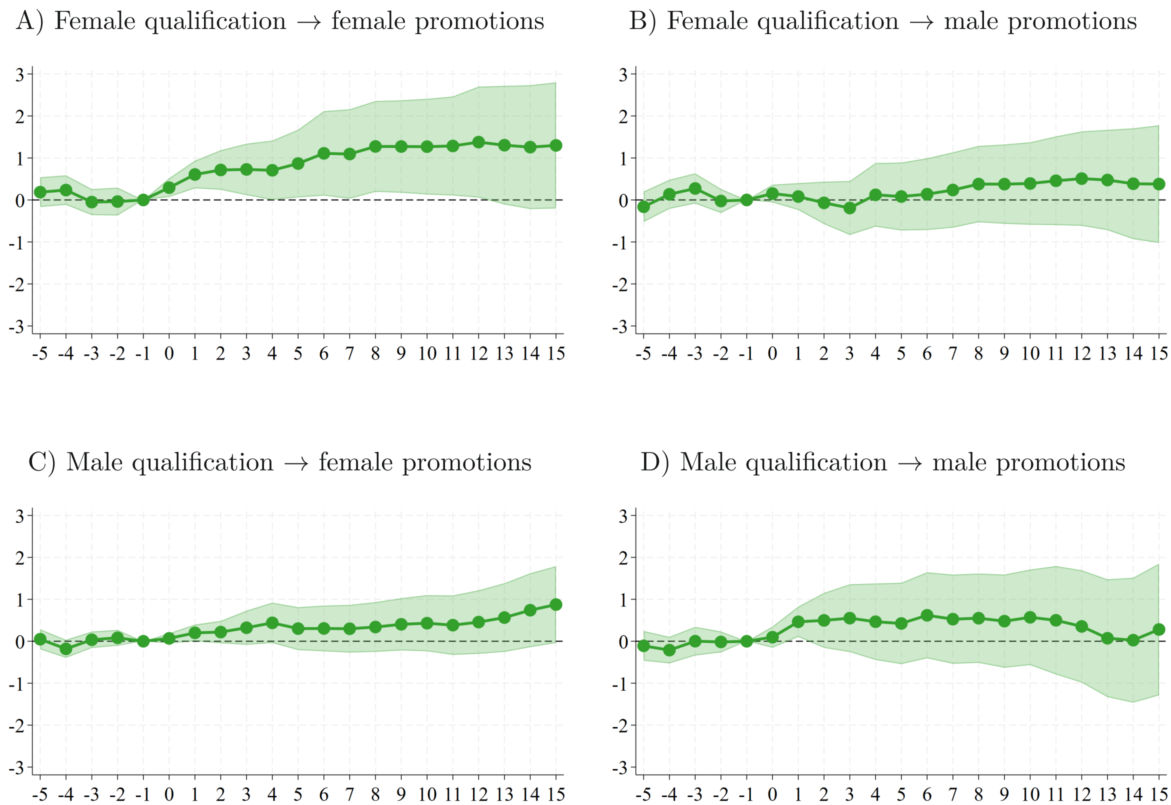
Figures

Figure 1: Individual-level effects of qualifications on promotions



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. The left column reports effects on promotion to Associate Professor and the right column effects on promotion to Full Professor. Panels A and B pool all candidates; Panels C and D report effects for women, and Panels E and F for men. The unit of observation is the candidate–exam cell. All specifications control for the candidate's research output at the time of the exam and for whether the candidate already held a tenured position, and include exam and qualification-year fixed effects. Standard errors are clustered at the exam level. Shaded areas represent 95% confidence intervals.

Figure 2: Department-level effects of qualifications on promotions to Associate Professor

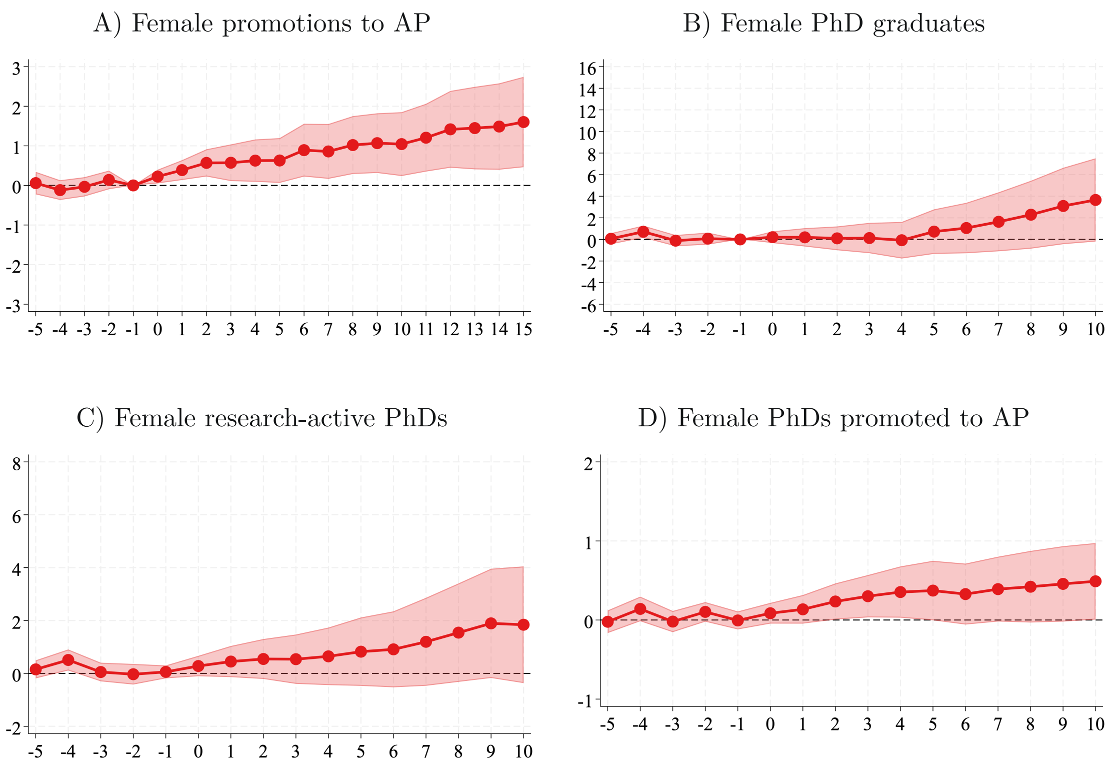


Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from equation (6). Panels A and B show the effect of an exogenous female qualification to *Associate Professor* on the number of female and male promotions to *Associate Professor*, respectively. Panels C and D show the corresponding 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 control for the number of applicants, pre-exam department faculty size, pre-exam PhD cohort size, and promotion rates in the years preceding the exam, and include exam and qualification-year fixed effects. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure 3: Department-level effects of qualifications to Associate Professor on 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 department's PhD graduates. Each column corresponds to one outcome: the cumulative number of PhD graduates, the number who remain research-active five or more years after graduation, and the number promoted to Associate Professor within ten years of graduation. The label above each row of panels indicates the combination of the qualifying candidate's gender and the students' gender. 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 control for the number of applicants, pre-exam department faculty size, pre-exam PhD cohort size, and promotion rates in the years preceding the exam, and include exam and qualification-year fixed effects. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Figure 4: Department-level effects of qualifications in female-intensive subfields



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 subfields 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 subfields 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 control for the number of applicants, pre-exam department faculty size, pre-exam PhD cohort size, and promotion rates in the years preceding the exam, and include exam and qualification-year fixed effects. Standard errors are clustered at the department level. Shaded areas represent 95% confidence intervals.

Tables

Table 1: Candidate characteristics

	All	FP exams	AP exams
Characteristics at the time of the exam			
Female	0.34	0.27	0.40
Tenured	0.49	1.00	0.10
Any connection	0.32	0.35	0.29
Colleague	0.30	0.32	0.27
PhD advisor	0.03	0.03	0.02
Coauthor	0.05	0.07	0.04
Number of connections	0	1	0
Papers	24	33	17
Papers in Web of Science	12	17	8
Books	2	3	1
Book chapters	5	7	3
Total AIS	13.26	17.14	10.15
Average AIS	0.85	0.83	0.87
Outcomes			
Qualified	0.11	0.11	0.12
Promoted within 2 years	0.90	0.95	0.87
Promoted within 10 years	0.95	0.98	0.93
Promoted in the same department	0.84	0.96	0.76
Research active after 10 years	0.88	0.90	0.86
Not Qualified	0.89	0.89	0.88
Promoted within 2 years	0.17	0.15	0.17
Promoted within 10 years	0.54	0.58	0.50
Promoted in the same department	0.77	0.89	0.67
Reapplied in a later call	0.46	0.53	0.40
Research active after 10 years	0.81	0.84	0.78
Number of applications	31,243	13,444	17,799

Notes: The unit of observation is an application to a qualification exam. Column (1) pools both ranks; columns (2) and (3) report Full Professor and Associate Professor exams separately. A connection is coded when an evaluator was the candidate's PhD advisor, a coauthor, or a colleague at the same university at the time of the exam; *Number of connections* is the count of such ties among the seven committee members. Research records are measured at the time of the exam and reported in levels. The outcomes are reported separately for candidates who qualified and for those who did not. *Promoted in the same department* is computed among candidates who were promoted. *Reapplied in a later call* indicates that the candidate applied again in a later *habilitación* call. *Research active after 10 years* indicates at least one journal article, book, or book chapter in the three years around the tenth year after the exam. Appendix Table A1 reports the same characteristics separately for women and men.

Table 2: Department characteristics

	All		AP exams		FP exams	
	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Before the exam						
Associate Professors	8.93	9.96	8.31	9.97	9.58	9.92
Women	3.20	4.25	2.92	4.13	3.50	4.35
Full Professors	2.81	3.54	2.66	3.56	2.96	3.51
Women	0.40	0.81	0.37	0.77	0.42	0.85
Panel B. Exam results						
Candidates in the exam	2.21	2.09	2.35	2.36	2.05	1.77
Women	0.75	1.10	0.93	1.25	0.56	0.87
Qualified candidates	0.26	0.55	0.29	0.61	0.22	0.48
Women	0.08	0.31	0.11	0.36	0.05	0.23
Panel C. Over the ten years after the exam						
Promotions to Associate Professor	3.60	4.74	3.85	4.91	3.34	4.55
Women	1.42	2.14	1.49	2.17	1.36	2.10
Promotions to Full Professor	2.12	2.65	1.81	2.47	2.45	2.78
Women	0.59	1.13	0.48	1.00	0.71	1.24
PhD graduates	28.05	29.69	27.09	29.11	29.04	30.25
Women	13.46	15.80	12.82	15.34	14.12	16.24
PhD graduates research-active	15.02	16.10	14.55	15.78	15.51	16.41
Women	7.23	8.58	6.92	8.32	7.56	8.84
PhD graduates promoted to Associate Professor	1.35	2.35	1.45	2.46	1.25	2.22
Women	0.50	0.99	0.53	1.05	0.46	0.94
Department–exam cells	13,218		6,733		6,485	
Departments	4,000		3,330		2,642	

Notes: The unit of observation is a department–exam cell. Columns (1)–(2) include all cells; columns (3)–(4) include cells with at least one applicant to an Associate Professor (AP) exam, and columns (5)–(6) cells with at least one applicant to a Full Professor (FP) exam, so a department with candidates in both types of exam appears in both groups. All entries are counts. Panel A reports faculty stocks in the year before the exam. Panel B refers to the department's applicants in that exam. Panel C reports totals over the ten years following the exam. *PhD graduates research-active* counts graduates with at least one journal article, book, or book chapter five or more years after graduation; *PhD graduates promoted to Associate Professor* counts graduates promoted within ten years of graduation. Faculty stocks and promotions are reconstructed from the Spanish State Bulletin (BOE) and PhD graduations from TESEO, as described in Section 3.

Table 3: Gender differences in the impact of qualification failure

	Reapplication	Academic activity after		
	(1)	(2)	(3)	(4)
	Within habilitación	5 years	10 years	15 years
Failed	0.419***	-0.135	-0.153*	-0.076
	(0.075)	(0.087)	(0.092)	(0.099)
Failed × Female	-0.280**	-0.147	-0.207	-0.298*
	(0.120)	(0.136)	(0.157)	(0.153)
Outcome mean, successful men	0.00	0.89	0.87	0.85
Outcome mean, successful women	0.00	0.83	0.85	0.81
p-value: effect of failure for men	0.00	0.12	0.10	0.44
p-value: effect of failure for women	0.16	0.01	0.01	0.00
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 *habilitación* 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, whether the candidate already held a tenured position at the time of the exam, expected committee connection intensity, exam fixed effects, and year fixed effects. *Outcome mean* rows report raw sample means among successful candidates of each gender; these differ from the levels implied by the coefficients, which identify effects for candidates whose outcome was determined by the committee draw. The final two rows report *p*-values for the total effect of failure for men (*Failed*) and for women (*Failed* + *Failed* × *Female*). Standard errors, clustered at the exam level, are in parentheses.
*** $p < .01$, ** $p < .05$, * $p < .1$

Table 4: First-stage estimates, department level

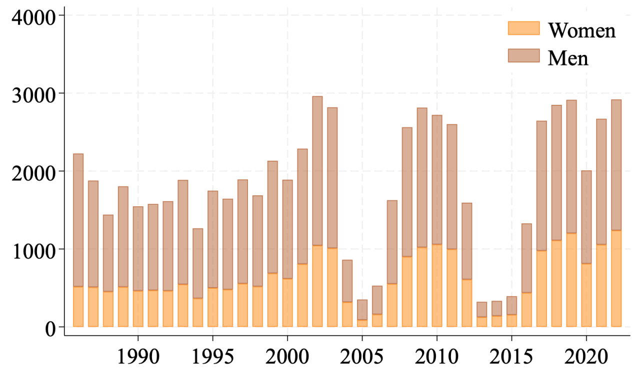
	Qualified	Qualified, Women	Qualified, Men	Qualified, Female-intensive	Qualified, Male-intensive
	(1)	(2)	(3)	(4)	(5)
Panel A. Associate Professor positions					
Connections	0.151***				
	(0.014)				
Connections, women		0.135***	0.020*		
		(0.020)	(0.011)		
Connections, men		0.004	0.143***		
		(0.006)	(0.018)		
Connections, female-intensive				0.130***	0.004
				(0.017)	(0.006)
Connections, male-intensive				0.031***	0.108***
				(0.010)	(0.019)
Observations	6732	6732	6732	6177	6177
KP rk Wald F-stat	117	34	34	19	19
SW F-stat	125	61	71	56	38
Panel B. Full Professor positions					
Connections	0.113***				
	(0.011)				
Connections, women		0.093***	0.017		
		(0.019)	(0.013)		
Connections, men		0.001	0.114***		
		(0.003)	(0.012)		
Connections, female-intensive				0.097***	0.013**
				(0.014)	(0.006)
Connections, male-intensive				0.008	0.091***
				(0.006)	(0.016)
Observations	6483	6483	6483	6161	6161
KP rk Wald F-stat	108	14	14	24	24
SW F-stat	118	31	91	50	38

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 female-intensive-area equivalent version of equation (5) described in Section 5.5. *Connections* measures the average deviation of realized from expected connectedness between candidates in a given department–exam and committee evaluators, with female-, male-, female-intensive- and male-intensive-specific variants defined analogously for the respective subgroups. All specifications control for expected connectedness, the number of applicants, pre-exam department faculty size, pre-exam PhD cohort size, and promotion rates in the years preceding the exam, and include exam and qualification-year fixed effects. Standard errors are clustered at the department level. The last two rows report the Kleibergen–Paap rk Wald F-statistic and the Sanderson–Windmeijer first-stage F-statistic for each endogenous regressor.
*** p < .01, ** p < .05, * p < .1

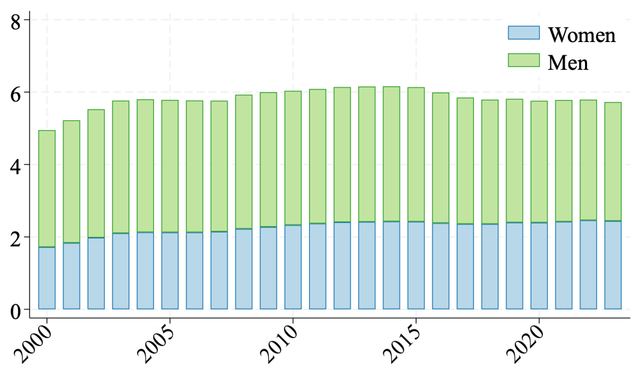
Appendix A — Descriptive statistics

Figure A1: Promotion Flows and Average Department Size by Rank and Gender

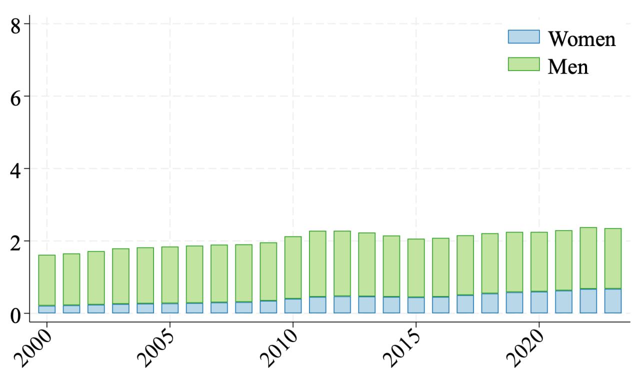
A) Promotion Flows



B) Associate Professors

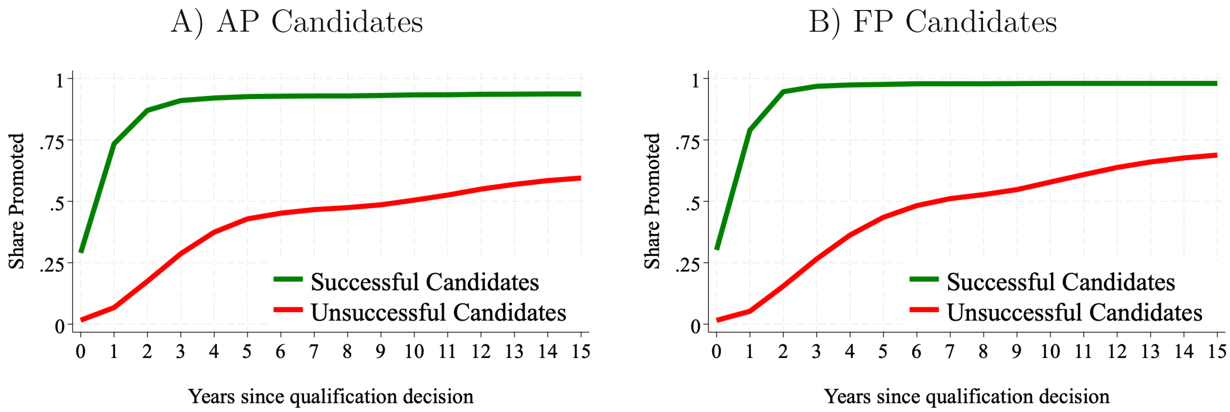


C) Full Professors



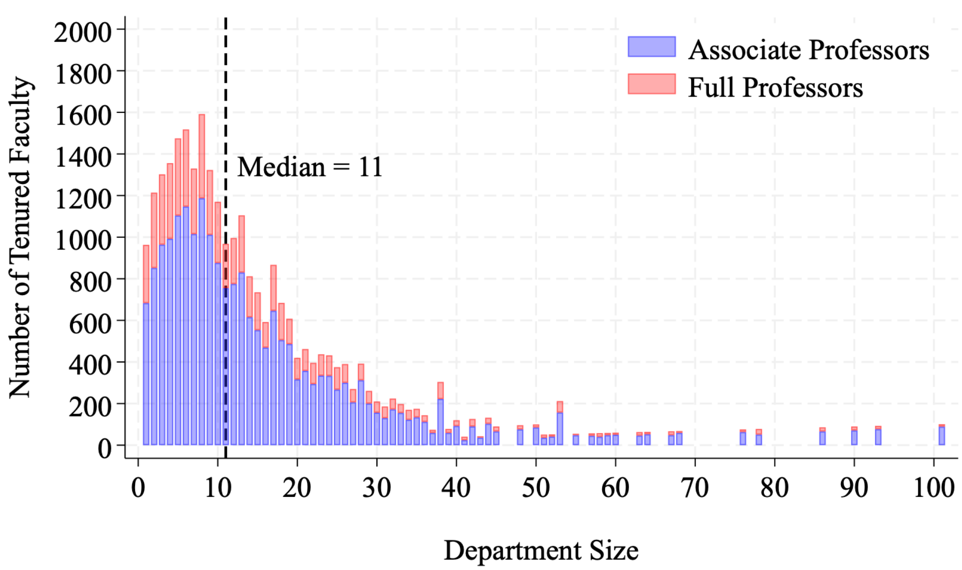
Notes: Panel A reports annual promotion flows, as recorded in the Spanish State Bulletin (BOE). Panels B and C report the average number of Associate and Full Professors per department, which we reconstruct from the same records following the procedure described in Section 3. All outcomes are reported separately for women and men.

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



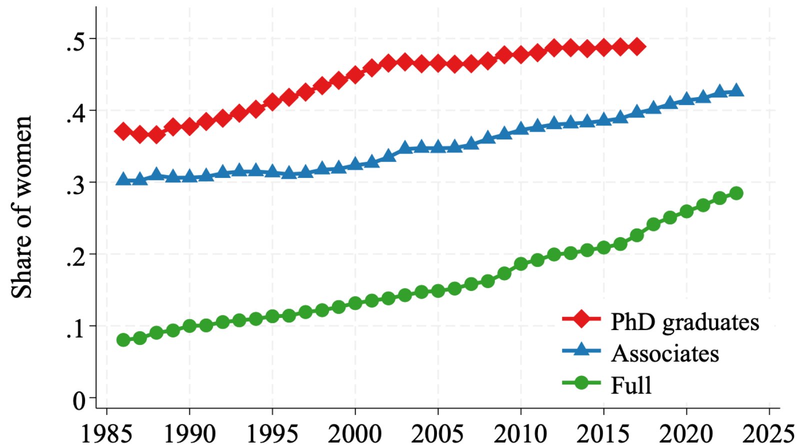
Notes: The figure plots the cumulative share of applicants who have been promoted, separately by qualification outcome, for candidates applying to Associate Professor positions (Panel A) and to Full Professor positions (Panel B). Outcomes are measured relative to the year of the qualification exam.

Figure A3: Distribution of tenured faculty by department size



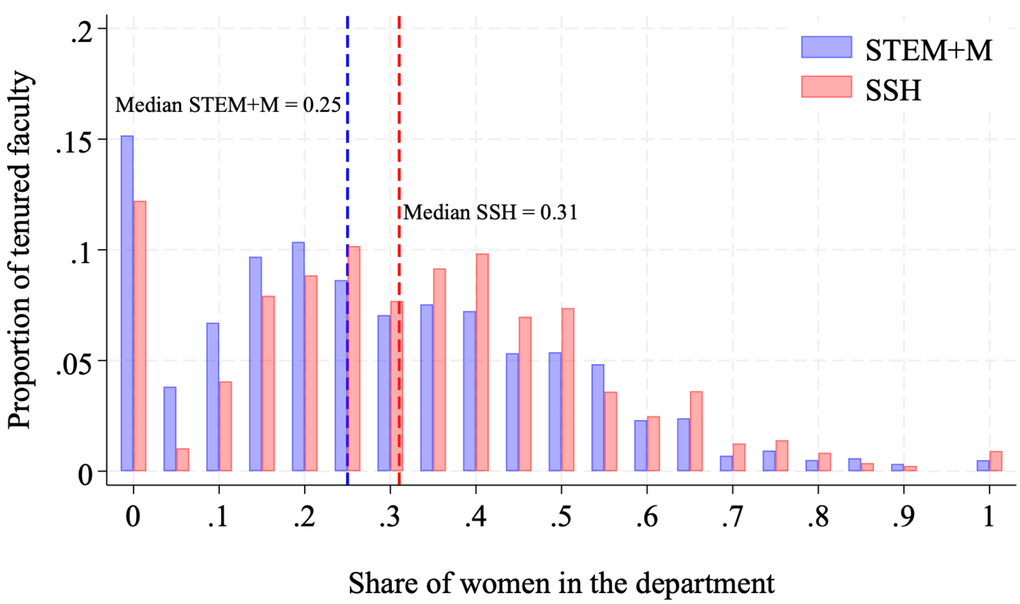
Notes: The figure shows the distribution of tenured faculty across departments in 2000. Department size is defined as the total number of Associate and Full Professors. Each stacked bar reports the number of tenured faculty employed in departments of a given size, separately by academic rank. The vertical dashed line marks the faculty-weighted median department size. 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



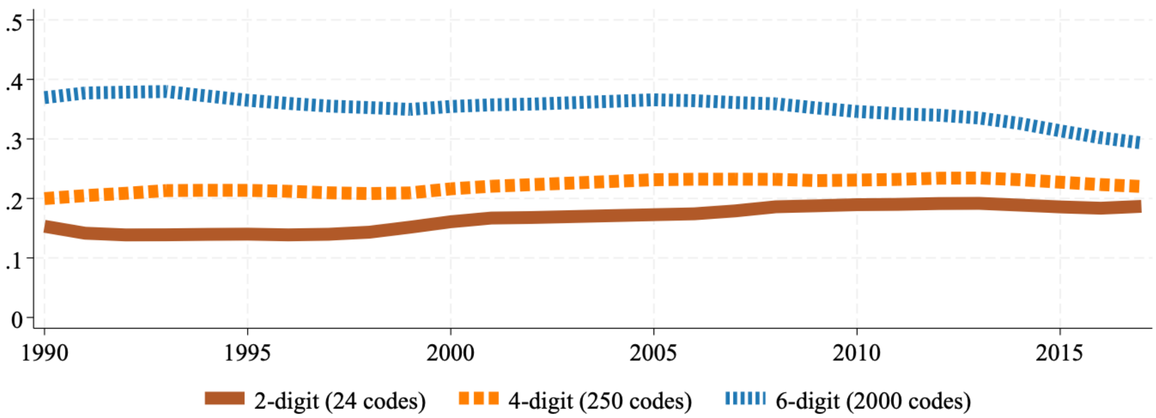
Notes: Each series reports, for each year, the share of women among that year's PhD graduates (1986–2017) and among all Associate and Full Professors (1986–2023). PhD shares are computed from dissertation records; faculty shares are our own calculations, based on the rosters of faculty eligible to serve on qualification committees and on promotion records from the Spanish State Bulletin (BOE).

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



Notes: The figure shows the distribution of tenured faculty by the share of women in their department in 2000, separately for STEM and medicine (STEM+M) and social sciences and humanities (SSH). Departmental shares are calculated among Associate and Full Professors and grouped into five-percentage-point bins. The dashed lines mark the faculty-weighted median in each field: 0.25 for STEM+M and 0.30 for SSH. 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 the TESEO dataset.

Table A1: Candidate characteristics, by gender

	Overall (1)	Men (2)	Women (3)	Women–Men (4)	SE (5)
Panel A. Applications for Full Professor [N= 13,444]					
Female Candidate	0.27				
Tenured	1.00	1.00	1.00	0.00	0.00
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 connection	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
Connections	0.573	0.585	0.543	-0.042**	0.019
Panel B. Applications for Associate Professor [N= 17,799]					
Female Candidate	0.40				
Tenured	0.10	0.11	0.10	-0.01***	0.00
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 connection	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
Connections	0.438	0.449	0.423	-0.026**	0.013

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: Candidate characteristics, by connection status and gender

	Overall sample			Women			Men		
	Connected	Unconnected	Diff.	Connected	Unconnected	Diff.	Connected	Unconnected	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A									
All candidates									
Qualified	0.18	0.09	0.09***	0.16	0.08	0.08***	0.18	0.09	0.10***
Papers	0.01	-0.00	0.02	-0.13	-0.16	0.03*	0.08	0.08	0.00
Papers in Web of Science	0.01	-0.00	0.01	-0.08	-0.09	0.01	0.05	0.04	0.00
Books	-0.00	0.00	-0.00	-0.10	-0.10	-0.00	0.05	0.05	-0.00
Chapters	-0.01	0.01	-0.02*	-0.10	-0.03	-0.06***	0.03	0.03	0.00
Total AIS	0.02	-0.01	0.02**	-0.07	-0.08	0.01	0.06	0.03	0.03*
Average AIS	0.01	-0.01	0.02	-0.03	-0.03	-0.01	0.03	0.00	0.03*
Citations	0.02	-0.01	0.04***	-0.06	-0.09	0.03*	0.06	0.03	0.03**
Number of observations	9911	21332		3294	7428		6617	13904	
Panel B									
Successful candidates									
Papers	0.31	0.38	-0.07**	0.12	0.16	-0.04	0.39	0.50	-0.10**
Papers in Web of Science	0.26	0.33	-0.08**	0.10	0.17	-0.07	0.32	0.42	-0.09**
Books	0.09	0.06	0.03	0.01	-0.01	0.01	0.13	0.10	0.03
Chapters	0.17	0.20	-0.03	0.12	0.17	-0.06	0.20	0.22	-0.02
Total AIS	0.23	0.31	-0.08**	0.08	0.16	-0.08	0.30	0.39	-0.09**
Average AIS	0.12	0.20	-0.08**	0.09	0.17	-0.08	0.13	0.21	-0.08*
Citations	0.25	0.34	-0.09**	0.13	0.17	-0.04	0.30	0.42	-0.12**
Number of observations	1744	1829		530	613		1214	1216	
Panel C									
Unsuccessful candidates									
Papers	-0.05	-0.04	-0.01	-0.17	-0.18	0.01	0.01	0.04	-0.03*
Papers in Web of Science	-0.05	-0.03	-0.01	-0.12	-0.11	-0.00	-0.01	0.01	-0.02
Books	-0.02	-0.01	-0.01	-0.12	-0.11	-0.01	0.03	0.05	-0.02
Chapters	-0.05	-0.01	-0.04***	-0.14	-0.05	-0.09***	-0.01	0.01	-0.02
Total AIS	-0.03	-0.04	0.01	-0.10	-0.11	0.01	0.01	-0.00	0.01
Average AIS	-0.01	-0.03	0.01	-0.06	-0.04	-0.01	0.01	-0.02	0.03
Citations	-0.02	-0.04	0.02*	-0.09	-0.11	0.02	0.01	-0.01	0.02
Number of observations	8167	19503		2764	6815		5403	12688	

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 tie to the candidate (PhD advising, coauthorship, or shared affiliation). Panel A includes all applicants, Panel B includes successful candidates, and Panel C includes unsuccessful candidates. Columns (1)–(3) report results for the overall sample, columns (4)–(6) for women, and columns (7)–(9) for men. Within each group, the first two columns report mean productivity measures for connected and unconnected candidates, respectively, standardized within exam, and the third column reports the difference between connected and unconnected candidates.
*** $p < .01$, ** $p < .05$, * $p < .1$

Table A3: Characteristics of PhD graduates in Spain

	Overall (1)	Men (2)	Women (3)	Women–Men (4)
Female graduate	0.47			
Female advisor	0.21	0.16	0.27	0.11***
Years to Graduate	5.54	5.50	5.59	0.10***
Staying in Academia	0.53	0.54	0.52	-0.02***
Promoted to AP	0.10	0.12	0.08	-0.04***

Notes: The unit of observation is a PhD graduate from a Spanish public university between 1990 and 2017; the sample comprises 203,542 graduates. Column (1) reports full-sample means. Columns (2) and (3) report means for men and women, respectively. Column (4) reports the difference in means (women minus men), with stars indicating statistical significance. *Years to Graduate* is available for the 30% of graduates whose PhD starting year is observed.
*** $p < .01$, ** $p < .05$, * $p < .1$

Appendix B — Randomization Procedure and 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.[22] The selection was carried out by Ministry officials using a drum

22. 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?docum

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 serve on a given committee.[23] 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

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

entId=0901e72b8008d9f6).

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

The expected number of connections is then:

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

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

Associate Professor exams. In Associate Professor exams, three evaluators are drawn from the pool of eligible Full Professors and four from the pool of eligible Associate Professors. We compute the probabilities p_R^{FP}, p_E^{FP}, p_R^{AP}, and p_E^{AP} analogously to the Full Professor case, separately for each pool. The expected number of connections is:

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

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 — The Timing of Tenure and Family Formation in Spanish Academia

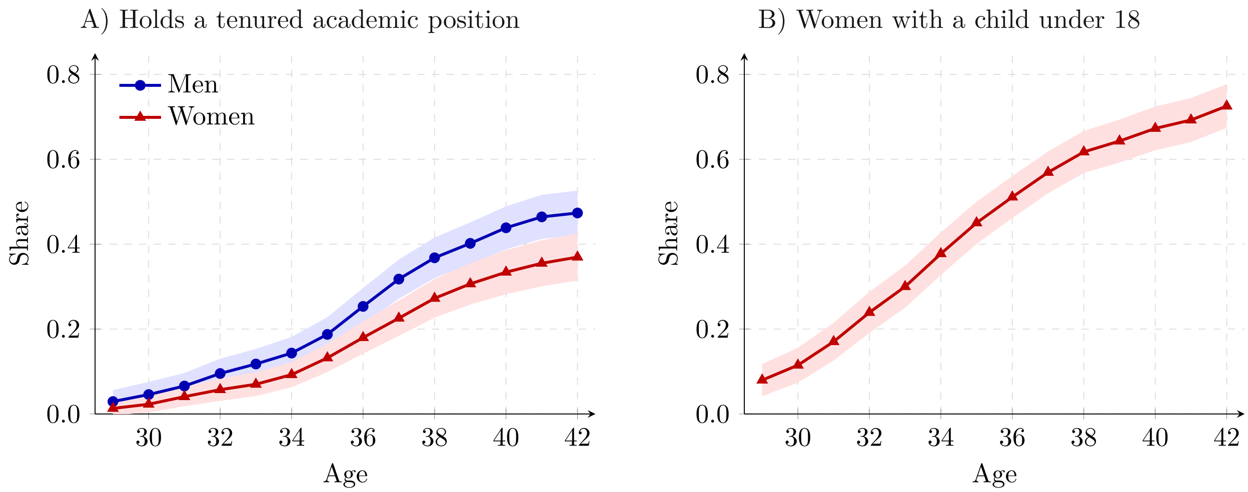
Two features of academic careers in Spain bear on the interpretation of our results. We document them using the *Encuesta sobre Recursos Humanos en Ciencia y Tecnología*, a survey of doctorate holders conducted by the Spanish National Statistics Institute in 2006 and 2009. We pool the two waves and restrict attention to respondents who completed their doctorate before age 35, so that career stage is broadly comparable across ages.

Panel A of Appendix Figure C1 reports the share holding a tenured university position, by age and gender. The share rises steeply through the late thirties for both, but women are less likely to hold one at every age: 22% against 34% at age 37, and 33% against 45% at age 40. The gap opens in the early thirties and does not close within the observed range. Panel B reports the share of women with a child under 18. It is only 10% at age 30, crosses one half at age 36, and reaches 73% by age 42.

Read together, the two panels describe a leaky pipeline in which the qualification decision arrives at a point when family responsibilities are already in place for most women, rather than before them. Both panels are cross-sectional associations from a separate data source and are not causal; they describe the setting in which the promotion decisions we study take place. Panel B counts only children under 18 living with the respondent, so it understates parenthood at older ages.

Our records do not include candidates' dates of birth, but their age at application can be approximated in two steps. Linking candidates to their dissertations in TESEO, the median candidate applies four years after completing the doctorate at the Associate Professor level and fifteen years after at the Full Professor level. According to the survey, doctorates completed in Spain over the corresponding periods were obtained at a median age of 33 for men and 31 for women. Together, these place the median Associate Professor candidate in the mid-thirties and the median Full Professor candidate in the mid-forties, assuming that applicants are not selected on age within a doctoral cohort.

Figure C1: Tenure and family formation among Spanish doctorate holders



Notes: The figure uses the *Encuesta sobre Recursos Humanos en Ciencia y Tecnología* (Instituto Nacional de Estadística), pooling the 2006 and 2009 waves and restricting to respondents who obtained their doctorate before age 35. Panel A reports the share of all doctorate holders who hold a tenured university position (*Profesor Titular* or *Catedrático*), separately for women and men; respondents working outside higher education are included and counted as not tenured. Panel B reports the share of female doctorate holders with at least one child under 18 in their care, irrespective of marital status. Lines are three-year moving averages; shaded areas represent 95% confidence intervals. Panel A is based on 8,813 respondents and Panel B on 4,398.

Appendix D — Randomization Check and First-Stage Estimates

This appendix reports diagnostic evidence supporting the identification strategy: balance tables at the individual and department level, followed by first-stage estimates at both levels.

Table D1: Balance under random committee assignment, individual level

Panel A.	Not controlling for expected connections			
	Papers	Books	Chapters	AIS
	(1)	(2)	(3)	(4)
Connections	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)
Connections	-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 connections 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 D2: Balance under random committee assignment, department level

Panel A.	Not controlling for expected connections				
	Candidates	Papers (z)	AIS (z)	Chapters (z)	Books (z)
Connections	0.068**	0.114***	0.093***	0.082***	0.070***
	(0.031)	(0.016)	(0.017)	(0.015)	(0.016)
Panel B.	**Controlling for expected connections**				
	Candidates	Papers (z)	AIS (z)	Chapters (z)	Books (z)
Connections	-0.050*	-0.012	-0.012	-0.006	-0.001
	(0.028)	(0.014)	(0.015)	(0.014)	(0.014)
Observations	13,215	13,215	13,215	13,215	13,215

Notes: The table reports OLS estimates at the department–exam level (including exams at 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 connections 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 D3: 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 connections	0.605***	-0.444***	0.113***	0.090***	0.052**	0.053*
	(0.039)	(0.025)	(0.026)	(0.027)	(0.026)	(0.027)
Male connections	-0.246***	0.239***	0.074***	0.064***	0.070***	0.047***
	(0.014)	(0.026)	(0.017)	(0.019)	(0.017)	(0.017)
Joint F-test [p-value]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.006]
Panel B.	**Controlling for expected connections**					
	Female candidates	Male candidates	Papers (z)	AIS (z)	Chapters (z)	Books (z)
Female connections	-0.003	0.011	-0.016	-0.005	-0.058**	0.000
	(0.039)	(0.029)	(0.028)	(0.029)	(0.028)	(0.027)
Male connections	-0.017	-0.055**	-0.005	-0.012	0.022	0.003
	(0.012)	(0.025)	(0.018)	(0.019)	(0.018)	(0.019)
Joint F-test [p-value]	[0.372]	[0.085]	[0.787]	[0.791]	[0.086]	[0.984]
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 at 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 connections between female and male candidates and committee members. Panel A excludes controls for expected connections; Panel B includes them. All specifications control for the number of female and male Associate and Full Professors in the department one year prior to the exam, as well as examination and year fixed effects. The joint F-test tests whether both instruments are simultaneously zero. Standard errors are clustered at the department level.

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

Table D4: 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
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	127	97	19	129	132	97	52

Notes: The table reports first-stage estimates of equation (1), where the outcome is an indicator for qualification and the explanatory variable, *Connections*, is the deviation of the realized number of ties to committee members (PhD advising, coauthorship, or shared affiliation) from its expectation under random assignment. All specifications control for expected connectedness, exam fixed effects, and qualification-year fixed effects. Productivity variables are measured at the time of the exam and standardized within exam. 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

Appendix E — 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.

E.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 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:

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

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 advising, 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:

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

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

E.2 Comparison across specifications

Appendix Tables E1 and E2 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 E1: First-stage estimates, alternative instrument definitions

	Binary	Count	Separate	Intensity
Outcome: Qualified	(1)	(2)	(3)	(4)
At Least One Connection	0.088***			
	(0.009)			
Connections		0.063***		
		(0.006)		
PhD Advisors			0.124***	
			(0.023)	
Colleagues			0.048***	
			(0.006)	
Coauthors			0.095***	
			(0.021)	
Connection Intensity				0.064***
				(0.006)
Observations	17762	17762	17762	17762
KP rk Wald F-stat	105.260	98.104	48.160	132.190

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 connected evaluators; column (3) includes the count of connected evaluators, separately by tie type (shared affiliation, PhD advising, and coauthorship); and column (4) considers a composite index equal to the number of connection 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 E2: Second-stage estimates, alternative instrument definitions

	Binary	Count	Separate	Intensity
Outcome: Promotion within 3 yrs	(1)	(2)	(3)	(4)
Qualified Candidate	0.506***	0.607***	0.582***	0.587***
	(0.114)	(0.118)	(0.086)	(0.093)
Observations	17762	17762	17762	17762

Notes: The table reports 2SLS estimates using the alternative instrument definitions described in Table E1. 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 F — Further results

This appendix reports additional 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. Next, we 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 orientation.

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 F4 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 F2 and F3 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 Heterogeneity by Department Composition and Field

Appendix Table F1 reports the estimates discussed in Section 5.4, by baseline female representation among senior faculty (Panel A) and by broad field group (Panel B).

Table F1: 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 AP (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)
STEM+M	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 control for the number of applicants, pre-exam department faculty size, pre-exam PhD cohort size, and promotion rates in the years preceding the exam, and include exam and qualification-year fixed effects. 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 the social sciences and humanities, while *STEM+M* includes science, technology, engineering, mathematics, and medicine. Standard errors, clustered at the department level, are reported in parentheses.
*** $p < .01$, ** $p < .05$, * $p < .1$

F.3 Research Orientation

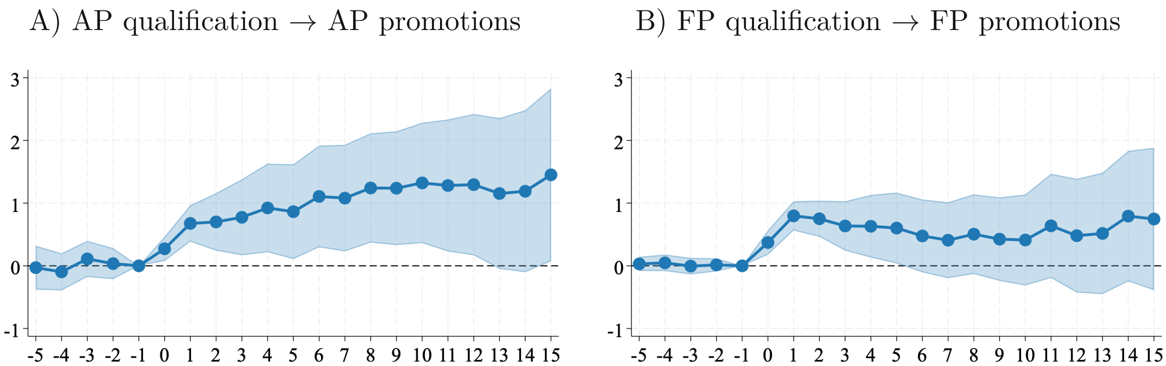
Appendix Table F2 reports the ten-year estimates underlying the main analysis of qualifications in female-intensive subfields. Panel A gives the effects on female outcomes plotted in Figure 4; Panel B gives the corresponding effects on male outcomes. Across male promotions, male PhD graduates, male research-active PhD graduates, and male PhD graduates later promoted to Associate Professor, the estimates are small and imprecise. The table also reports department-level effects of qualifying candidates working in male-intensive subfields. Unlike qualifications in female-intensive subfields, these shocks generate no clear effects on female or male promotions or PhD outcomes.

Table F2: Department-level effects of qualifications in female- and male-intensive subfields, by gender of outcome

	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				
Qualified, female-intensive	1.044**	3.656*	1.842	0.490**
	(0.411)	(1.967)	(1.126)	(0.247)
Qualified, male-intensive	0.027	-2.186	-0.775	-0.052
	(0.545)	(2.787)	(1.638)	(0.352)
Panel B. Outcomes of Men				
Qualified, female-intensive	0.796	1.600	1.204	0.173
	(0.507)	(1.917)	(1.124)	(0.297)
Qualified, male-intensive	0.126	-1.021	0.508	0.279
	(0.754)	(3.320)	(1.940)	(0.460)

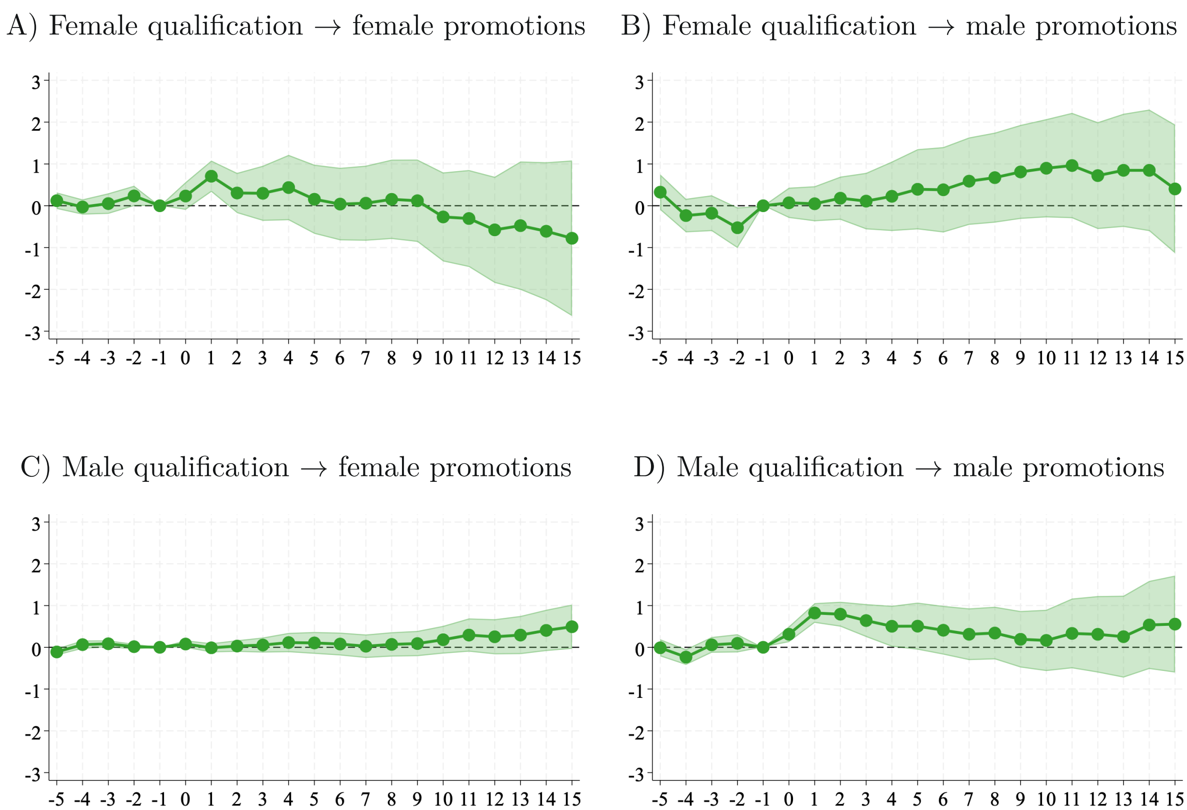
Notes: The table reports 2SLS estimates at the department–exam level for applications to Associate Professor positions. Outcomes are measured ten years after the qualification exam. Panel A reports outcomes for women and Panel B for men. *Qualified, female-intensive* and *Qualified, male-intensive* are the number of a department's candidates qualifying in each type of subfield; female-intensive subfields are defined in Section 3. All specifications include controls listed in Section 4.2. Standard errors, clustered at the department level, are reported in parentheses.
*** p < .01, ** p < .05, * p < .1

Figure F1: Impact of exogenous qualifications on the department



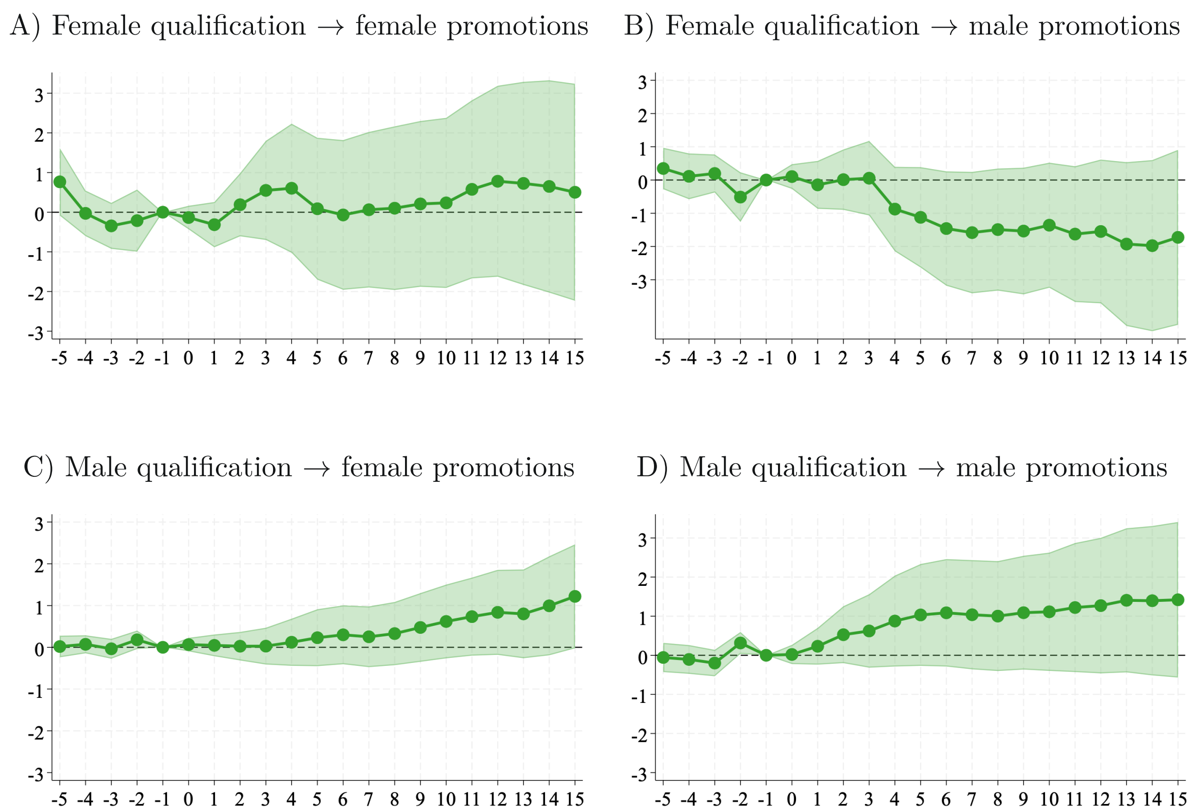
Notes: The figure shows 2SLS estimates of coefficient δ_1 from equation (4), capturing the impact of exogenously qualifying a candidate for Associate Professor on the cumulative number of promotions to Associate Professor (Panel A), and the impact of exogenously qualifying a candidate for Full Professor on the cumulative number of promotions to Full Professor (Panel B), over the years following the qualification exam. The unit of analysis is 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 F2: Department-level effects of female and male qualifications to *Full Professor* on promotions to *Full Professor*



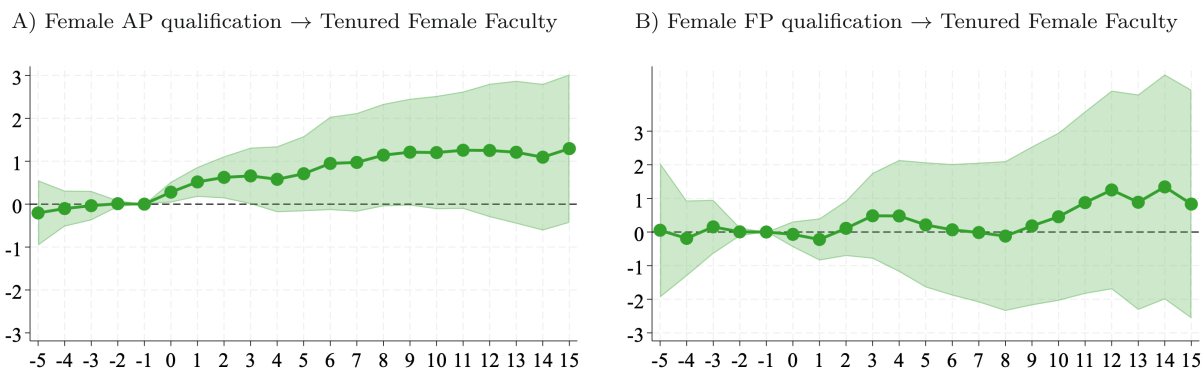
Notes: The figure reports 2SLS estimates of coefficients δ_1 and δ_2 from equation (6). Panels A and B show the effect of an exogenous female qualification to *Full Professor* on the cumulative number of female and male promotions to *Full Professor*, respectively. Panels C and D show the corresponding 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. Shaded areas represent 95% confidence intervals.

Figure F3: 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. Shaded areas represent 95% confidence intervals.

Figure F4: Department-level effects of a female qualification on tenured female faculty

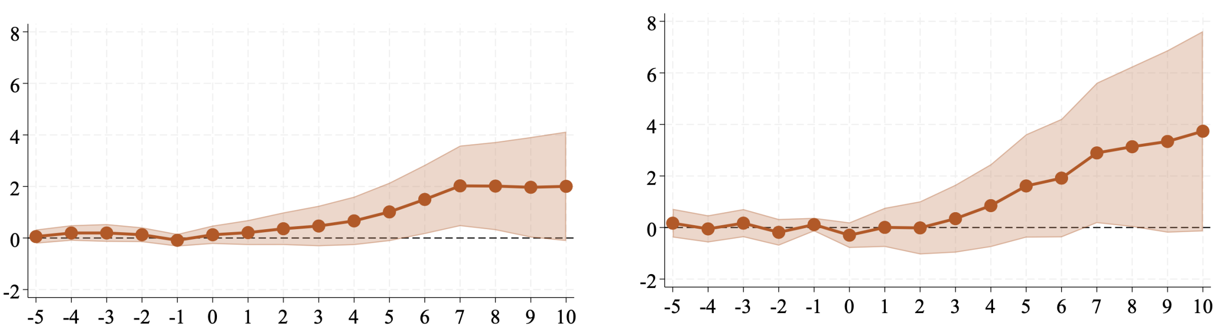


Notes: The figure reports 2SLS estimates of coefficient δ_1 from equation (6), capturing the impact of exogenously qualifying a woman on departmental faculty outcomes. Panel A shows the effect of qualifying a woman to Associate Professor on the stock of tenured female faculty. Panel B shows the effect 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. Shaded areas represent 95% confidence intervals.

Figure F5: 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



Notes: The figure reports 2SLS estimates of coefficient δ_1^f from equation (6), capturing the impact of exogenously qualifying a woman to *Associate Professor* on the cumulative number of female PhD graduates, split by the gender of the lead supervisor: graduates supervised by women (Panel A) and by men (Panel B). 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.

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% of Full Professors and 70% 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*).[24] 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.[25]

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.[26] We exclude 42 applications from eight candidates who reapplied after being promoted in a previous wave and were subsequently removed from the selection process. 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.

24. 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).

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

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

Scopus is a large-scale, multidisciplinary abstract and citation database launched by Elsevier in 2004, widely used to benchmark international academic output. It offers structured metadata on articles published in indexed journals, providing information on publication years, citation counts, journal-level indicators, and author affiliations. 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. 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 come 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). We use the version of these concepts, and the algorithm to assign them, provided by 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.[27] 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.

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 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 normalize 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.

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

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

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.

We then normalize 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 $p_{i,t}$.