



Gender imbalance in doctoral education: an analysis of the Spanish university system (1977–2021)

Rodrigo Sánchez-Jiménez¹ · Iuliana Botezan² · Jesús Barrasa-Rodríguez³ · Mari Carmen Suárez-Figueroa⁴ · Manuel Blázquez-Ochando⁵

Received: 9 September 2022 / Accepted: 23 January 2023 / Published online: 11 February 2023
© The Author(s) 2023

Abstract

Doctoral education is a key feature of university systems, as well as a basic foundation of scientific practice. That period culminates in a dissertation and examination of the candidate that has been studied from several points of view. This paper reports the results of an analysis on the evolution and characteristics of gender imbalance of a complete doctoral system for a wide period of time. Data from the database Teseo was used in order to identify the individuals involved in the process, the scientific fields in which the dissertations were classified, and the institutions in which the examination took place. Results: the Spanish system shows a clear evolution towards gender balance, but also some concerning trends that are worth tracking. Seemingly, STEM disciplines look to be evolving more slowly than other branches of science in several aspects. A leaky pipeline is characterized in this system around the roles of supervisors, candidates, members and chairs of the dissertation committees. Gender assortativity is also studied and described, and its possible effects discussed around the academic relations that surround doctoral examination.

Keywords Gender imbalance · STEM · Dissertations · Teseo · Leaky pipeline · Gender assortativity

✉ Rodrigo Sánchez-Jiménez
rodsanch@ucm.es

¹ Library and Information Science Department, SCImago Group, Universidad Complutense de Madrid, Madrid, Spain

² Library and Information Science Department (Internet Medialab Research Group), Universidad Complutense de Madrid, Madrid, Spain

³ Sales Engineering EMEA, Neo4j, London, UK

⁴ Ontology Engineering Group (Artificial Intelligence Department), Universidad Politécnica de Madrid, Madrid, Spain

⁵ Library and Information Science Department, Universidad Complutense de Madrid, Madrid, Spain

Introduction

Gender imbalance and gender bias in science have been studied and described for a long time. Zuckerman and Cole (1975) already described this issue in quantitative terms and advanced the “principle of triple penalty” (cultural inappropriateness/perceived incompetence/direct discrimination). Shiebinger (1987) went over the very abundant literature on the history of women in science and described how at least the number of female scientists was growing faster (low numbers having been associated with that cultural inappropriateness) but the perception of a lesser competence by women (women were systematically employed in less prestigious jobs in the academia) and blatant discrimination (unjustified salary gaps were huge) was ever persistent. Etzkowitz, et al (1992) moved on to put the focus on the de-genderization of science and society, and on the existence of “different gender styles of scientific work”, an idea that has been one way or another behind many studies comparing output, collaboration, and impact of men and women. Bordons et al. (2003) acknowledge this factor as a warning to interpret their SCI-based results but take it a step further. They also explained the cumulative advantage of achieving high ranks in academia over productivity, which in turn accounts for the gender differences in productivity. Several years later, Larivière et al., (2011) reached a somewhat different conclusion, finding again that gender differences were present in terms of production and funding, although the nature of these differences was complex. The subject is therefore very much open to debate, and the focus on its study has varied significantly over time (Tomassini, 2021).

The literature on the participation of women in science has grown steadily over the time, but the last 15 years have seen a truly remarkable increase in the attention paid by scholars (Larivière et al., 2013), (Holman et al., 2018) (Huang et al., 2020) and the impact of their research. The study of the role of women in scientific research and their contribution to scientific output (Abramo et al., 2009; Holman et al., 2018; Macaluso et al., 2016; West et al., 2013) has thus become a fairly common research topic in bibliometrics and research evaluation studies. Overall, the issue of the relative presence of women in science has been described as “progressing”, but still far from ideal. Huang et al. (2020) made a comparison on gender inequality/bias in scientific careers throughout history (1955–2010). The research was based on data extracted from the collection of Web of Science (WOS) databases about bibliographic references by country (83) and subjects (15). They found that over the 55-year interval they analyzed, only 27% of the authors of the WOS were women, growing from just a 12% in 1955, to a 35% in 2005, with a very unequal distribution by country: from a 28% in Germany to parity (50%) in Russia. Larivière et al. (2013) came to very similar conclusions on their study using the same source of data: A 70–30% split between men and women authors.

One of the main axes of study has been the relevance of gender in explaining the social relationships inherent to science, and specifically collaboration between authors. This kind of studies have taken a more powerful perspective when a network study was used (Araújo & Fontainha, 2017; Dehdarirad & Nasini, 2017). The issue of whether women tended to collaborate more with women than they collaborate with men (homophily or gender assortativity), and to what extent has also been an interesting by-product of this kind of studies (Jadidi et al., 2018). As a result, it seems clear that academics tend to collaborate more with other academics of the same gender, and the phenomenon seems to be increasing, at least regarding publishing (Holman & Morandin, 2018).

Over the years, preoccupation with gender disparities has grown in specific scientific domains deemed more prone to gender inequality, for example STEM (Science, Technology, Engineering and Mathematics). Blackburn, (2017) offers a good overview on the subject covering all stages and every major point of interest. The high variability of gender imbalance over scientific domains (Holman et al., 2018) has been a relevant topic of study and appeared as a recurrent variable in the literature, having been described as horizontal segregation (Tomassini, 2021). Some of these topics have also been studied in the context of doctoral education, a secondary although relevant axis of research on science-making, career development and the sociological factors of science.

Doctoral education and gender bias

The path to becoming a doctor involves both the production of scientific output and a process of examination that requires the participation of a number of academics carrying out different roles. PhD candidates in Spain are expected to undergo a formal examination of their dissertation in a public act that is usually referred to as “doctoral dissertation defense”. The examination committee is usually formed by 5 members, although the number has varied over the years and universities. One of the members acts as chair and directs the process of examination. The chair/president is usually the most senior academic and with the highest academic degree. The composition of the committee is usually set by the department or faculty responsible for the doctoral program and superseded by other centralized instances of the university. In practice, the supervisor plays a very significant role in defining the members of the committee. This has led to the study of supervisor’s choosing of committee members as a structural feature (Delgado López-Cózar et al, 2006).

Supervisors, candidates, commission chairs and commission members concur around the thesis examination, but also before and after the candidates become doctors. The role of supervisors is clearly very influential in the candidate’s scientific career, both in terms of “learning the craft” and in introducing them to their academic networks (Wang et al., 2021). The members of the committee have an obviously important role in the outcome of the examination, but they also play a part in the process of creating scientific networks for the candidates (Breimer, 2013). The study of the relationship between committee networks and coauthorship networks has been the subject of recent research by (Bès et al., 2021), that describes the collaboration of candidates and committee members, finding that (although disciplinary difference exists) “committees stand (even defectively) not only for networks of interpersonal relationships achieved through the participation in an academic ritual of integration, but also for the working communities built up by disciplines”. The study of the examining committees has thus the potential of revealing deep levels of scientific collaboration. Both Olmeda et al. (2009) and Gonzalez-Alcaide & Gonzalez-Teruel (2020) elaborate on the relevance of the idea of the invisible college proposed by Crane (1972) when applied to the analysis of the social structures that might be detected in examination committees. This idea is also present in a less evident fashion in Repiso-Caballero, Torres-Salinas & Delgado-López-Cózar (2011) and López-Yepes (2002), that also explored the study of committees as social structures of science. Gender assortativity might play a role in these relations, and has also been studied for doctoral dissertations (Bu et al., 2020; Schluter, (2018));, although it is different from research collaboration in that it is intrinsically asymmetric, as a byproduct of the supervisor-candidate relationship.

The issue of the gender of candidates and supervisors is not only subject to study from the homophily point of view and its consequences in terms of collaboration. It is also studied as a cause for drop-out in the postdoctoral period (Gaule & Piacentini, 2018). Postdoctoral scientists, and the gap between the number of successful candidates and the number of actual scientists, is another very interesting aspect of the doctoral system. It has also been studied from a gender bias perspective (Borrego et al., 2010; Reybold et al., 2012).

Doctoral dissertations have been also studied as a genre of scientific output, with its own characteristics. Although most of the studies on scientific production are based on journal articles, widening the spectrum of analysis can also be beneficial, as dissertations offer a parallel viewpoint that is subject to a rigorous academic scrutiny (Delgado López-Cózar et al., 2006) and might thus provide us with a relevant indicator of scientific activity growth (Fernández-Cano et al., 2012). According to this, dissertations can be used as a means to describe thematic distribution and shift over time, much as regular journal output is used. A good number of studies have characterized the thematic evolution of scientific activity using PhD dissertations. To name but a few, Sugimoto's (2011) depiction of the evolution of research topics in Library and Information Science (LIS), (Ying & Xiao, 2012) analysis of the thematic structure of the field of tourism, (Zong et al., 2013) co-word study of LIS in China or the work of Delgado, Repiso and Torres-Salinas in different areas (Delgado López-Cózar et al., 2006; Repiso Caballero et al., 2012; Repiso et al., 2011). One interesting consequence of this kind of grey literature is the development over the years of a (Breimer & Mikhailidis, 2020) "thesis by published papers" mode, making its way even into the Social Sciences and Humanities, a traditional stronghold of the monograph thesis. Gender bias has also been studied in relation to the production of doctoral dissertations, the scientific and academic networks interweaven in the examination, and as an early stage of scientific careers.

Our closest precedent is the work of Villarroya et al., (2008), that studies the gender distribution of doctoral candidates, supervisors and examination committee members. This study used a sample of over 1.000 doctoral theses that spanned from 1990 to 2004 and covered all the major research areas. It largely relied on manual procedures for gender detection over the records of the Spanish database Teseo. It offered insights on the presence of women in the doctoral system and its evolution, the diversity of the distribution over different disciplines, the inequalities in certain roles during the doctoral process or the specific dynamics of collaboration as influenced by gender. Estela Hernández-Martín and colleagues also found heavy inequalities for the Universidad Politécnica de Madrid (UPM) data in the more recent period of 2006–2016 (Hernández-Martín et al., 2019): the ratio of female to male PhD students was 23/77 (female/male, %), and the gender imbalance among the thesis advisors was 15% women vs. 82% men, and 3% mixed, on average. In addition, gender imbalance was present in PhD committees, with a repeated predominance of male members. Many other studies have been published, although they usually focus very narrow subjects. Diaz-Kope et al., (2019) analyzed public administration, policy, and public affairs, Welsh and Abramson (2018) engineering, Vallejo et al. (2016) mathematics education, and (Castelló-Cogollos et al., 2015) sociology, to name but a few.

Our work tries to provide a broad view of the main issues that have been studied in the past by other researchers in relation to the gender gap in science at one of the earliest stages of the researchers' careers. In order to focus on the less studied doctoral period we have extracted data from the Spanish database Teseo, that keeps a systematic record of PhD dissertations successfully defended in Spain from 1977 onwards. Although the database has some coverage issues (Fuentes Pujol et al., 2010) as well as data quality problems (Sánchez-Jiménez et al., 2017), it has been successfully used as an information source for

meta-analysis and systematic reviews (Castellano et al., 2014; Catalá-López et al., 2012). It has also been abundantly used in the study of diverse fields through their dissertation literature, such as psychology (Musi-Lechuga et al., 2009), education (Curiel-Marín & Fernández-Cano, 2015), emergency medicine (Fernández-Guerrero, 2015), scientific medical information (Fernández-Guerrero et al., 2020) or sports science (Hernández-González et al., 2020). The scope of our research involves a very long period of time, from the late seventies to 2021, and a very wide range of scientific disciplines, that basically covers the whole spectrum of science. Our objective was thus to offer an update, but also to broaden our knowledge on a complete national doctoral system, and in some cases, offer some extra detail that was not available before.

Data and methods

Our data source (Teseo) is maintained by the Spanish Ministry of Universities. It includes information about PhD candidates (successful ones), advisors and members of the committees, as well as data from the dissertation itself (including a subject classification and abstract) and data regarding the home institution of the PhD program. We have heavily relied on the previous experience by (Sánchez-Jiménez et al., 2017) using a fairly similar data extraction process. A web scraper was put in place and its set-up was very much the same. We devised a simple algorithm to extract names and surnames from incorrectly processed name labels in order to improve authority control. After that, we used lists of female and male names (with their frequencies) in order to assign a gender to the individuals involved on the whole Spanish doctoral system. These two steps are described in more detail below.

Teseo lacks specific information related to PhD students who made their research work outside Spain and there are also dissertations that have not been included in the database for unknown reasons (approximately 10% of all dissertations) if we compare it to the most comprehensive source of information, Dialnet Tesis (<https://dialnet.unirioja.es/tesis>), which on the other hand lacks important information such as subject matters or committee composition. In addition, Teseo has a problem with homonymity (e.g., same name, middle name, paternal surname and maternal surname) and name variations, as well as spelling mistakes, which were also found.

We have primarily dealt with data on the academics involved in the process of defending and obtaining the PhD degree, and thus, most of our efforts involved cleaning, refining and unifying data on people. After downloading the dataset, which comprises information about more than 275.000 theses (from 1977 to September 2021), a grand total of 1.848.776 references to individuals was obtained. A breakdown of these according to their roles is shown on Table 1. One should note that although most of the thesis had a single supervisor,

Table 1 Roles of individuals mentioned in dissertation records

Role mentions	Frequency	Percentage
Candidate	275,211	14.89
Committee Member	1,198,268	64.84
Supervisor	375,307	20.3
Other	55	0.003
Total	1,848,776	100.00

many had two, and some had even three. The role and responsibility of supervisors was hard to establish for multi-supervised dissertations, as information was not consistent, so no difference was made between directors and co-directors.

As these mentions were completely oblivious of the idea of individuals, we had to try to match the names that appeared in different situations (mentions) in order to create a record for every person that was involved in the doctoral examination process. In other words, we had to try to perform a reasonable authority control procedure. Before we attempted to match mentions of individuals in different roles, we had to do some cleansing, consolidation and normalization on the labels that were used to identify those individuals.

In principle, data is inserted in the database following a systematic pattern; that is, family names first, and given names afterwards, separated by a comma. Nevertheless, 139.201 (7,5%) name labels had no division, or too many commas, and had to be divided algorithmically. In order to achieve this, an index of names and surnames was created using the tags of the mentions that had been correctly divided. Each doubtful label was analyzed in order to decide the best way to divide it in the two fields that we needed in order to identify individuals. The position of every particle of the label and the frequencies extracted from the index allowed us to narrow the amount of incorrect labels to a more discrete 13.518, or roughly 0.007% of all the references to people in the database. The procedure was tested with randomly selected labels (some 650) and yielded a 94.7% precision. We take this evaluation only as a rough estimate to work with but are not extremely worried about its correctness as the global amount of problematic labels is low.

After correcting the data, the different mentions that shared a common name and surname (exact match) were merged to a person's record. This provided us with slightly more than 430.000 records, identifying specific individuals that could be traced back to their appearances and roles in the original thesis records. Once this was accomplished, the problem of identifying gender in the names of the individuals was addressed.

We need to clarify that we did not attempt to identify or infer the gender of the individuals, but were merely trying to gather enough information to recognize names that are usually identified with female or male genders, as well as names that are used both by female and male individuals. This later case was specially challenging, as not only some names are used both for men and women (Cruz, Trinidad, Ventura, Hua, Yao) but some are used for females in some countries (Simone in France) and males in others (Simone in Italy). We used a fairly simple procedure based on the frequencies of the different names in association with one or more genders. Procedures like this have already been proved successfully before (Green et al., 2009; Karimi et al., 2016; Mislove et al., 2011; West et al., 2013). The data source used for this purpose was Spain's National Statistics Institute lists of female and male names with a frequency of over 20 (INE, 2021a).

After matching the names of individuals with names from this data source, 82% of the names were identified as Female, Male or Non-Binary. This last category included names that were used at least significantly (in 10% or more of the cases) by both genders. As shown in Table 2, the number of unrecognized names is still substantial, which might imply a performance problem regarding the method of gender identification. Jadidi et al. (2018) had reported significantly improved correctness in their assessment of the performance of several other gender recognition systems with Spanish scholars. We tested 500 random names with Genderize (<https://genderize.io/>), the top performing gender detection service for the Spanish data, as well as with our procedure based on INE's name data. After manually checking the 500 names, Genderize achieved 66,2% of correct guesses, while our method achieved 85,2%. We are well aware that there might be a bias in the names that are not correctly identified, as also noted by Huang et al., (2020), which found

Table 2 Gender of individuals mentioned in dissertation records

Estimated Gender	Authorities	Percentage
Female	131.231	30,09
Incorrect Format	8.405	1,93
Male	225.570	51,72
Not recognized	68.676	15,75
Non-binary	2.220	0,51
Total	436.102	100

the task of assessing the gender of individuals from some countries (China, Japan, Korea, Brazil, Malaysia, and Singapore) extremely difficult.

Results

Our work has been focused around four main aspects of the analysis of the gender gap in the doctoral examination process and its outcome. First of all, we analyzed the evolution of the percentage of women undertaking different roles in the process. After that, we studied the proportion of women according to the scientific domain of the dissertations, as well as the share of female participants according to the institutions in which the doctoral degree was obtained. Lastly, we looked at the problem of gender assortativity in the context of the committees, analyzing the assortativity of candidates and supervisors, and also the assortativity of supervisors and committee chairs.

Evolution of the gender gap in the doctoral system

We did know that the number of dissertations that had been defended successfully by female candidates had grown steadily and had been around 50% for the last few years (INE, 2021b). As the number of female doctoral graduates grew, the sheer number of females in the system was expected to increase, which indeed was the case, as Fig. 1 shows. Villarroya et al., (2008) had already reported an evolution from a 60%/40% in favor of men to a balanced situation in 2005, in terms of the gender of the candidates.

The number of individuals that had non-binary names has increased over the years, although data seems shaky during the first period. This might be related to the increase of foreign students, although it is an issue that would require further research.

The number of female candidates has grown to be equal (or even greater) than that of men for the last few years and has decreased slightly during the last years. It might be soon to see a clear tendency here, but this is an important feature of data. The percentage of female supervisors has increased greatly, having doubled from the year 2000, and increased fivefold from the early eighties. This pattern seems to have stopped growing for the last few years, as the percentage of female advisors has remained around 30% for the last 5 or 6 years. This is interesting (and worrisome) and has been analyzed below in relation with the trend of female candidates.

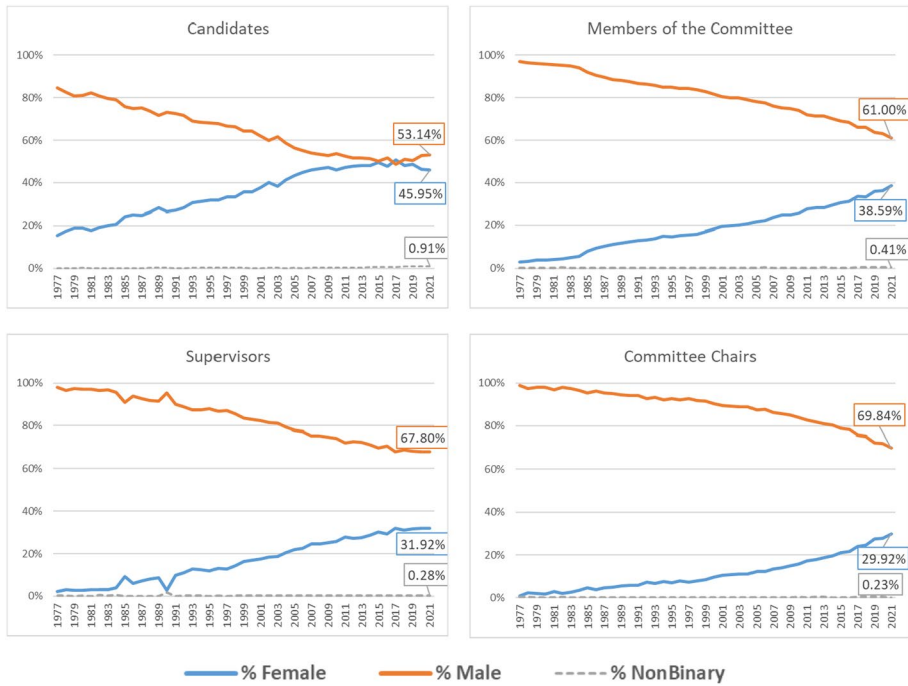


Fig. 1 Evolution of the gender of doctoral candidates, supervisors, committee chairs and members of the committee

PhD examination committee members have not been studied so amply in the literature, probably because of a lack of data, but also probably because of a lack of standardization of the roles participating in the committee, and the different shapes that those committees have adopted over the years. It was not strange to have committees of seven members, or committees chaired by the supervisor at the beginning of the data series, and it is not strange to see three committee members in data around the last few years. In any event, chairs do have a prominent role in these committees, while other roles (secretary, vocal) were identified less easily, and had a less evident meaning. We have tracked the evolution of both non chair members of the committee (independent of their roles) and chairs, which can be recognized more easily in data.

We were expecting to find a clear change of tendency around 2007, when new rules on the constitution of the committees were introduced. In the Spanish law on the effective equality of women and men (Government of Spain, 2007) there is an article (number 5, “Equality of treatment and opportunities in access to employment, training and promotion of professionals and in working conditions”) that affects the university environment for offering public employment for new doctors, together with article 52 (“Heads of management bodies”) that speaks of appointment to positions with management roles (of the General State Administration and their related public bodies) with a balanced presence between men and women.

The new rules de facto imposed gender balanced committees wherever possible. The number of committees consisting of 5 academics is hugely predominant, and even-number committees are most probably the consequence of faulty data or extraordinary

administrative situations. Considering this, gender balance could only be achieved as a whole, and not in particular cases. That is, the percentage of women that were taking part in the process of PhD examination should have increased significantly after that point, but it did not happen. Data shows a clear progression in the number of female members of the committees, but there are no bumps in the pattern. We can only hypothesize that this evolution might be originated by an organic process, that would probably be more tightly related to demographics, or changes in the academic culture.

The percentage of female committee members has grown steadily, and have not dwindled in the last years, as the supervisors' have. The last data show almost a 60/40% in favor of men, but it is relevant to point out that in the Spanish university, the percentage of women with a teaching contract rises to 42% (INE, 2021c), so in relation to the demographic base in the Spanish academia, growth might even be limited in the next years. The tendency towards reducing the gap is also clear in relation to committee chairs, but the gap in this role is still very important, being even greater than the male/female supervisor gap (70/30% vs a 68/32%, respectively). This bears a stark contrast with data from previous studies (Villaroya et al. 2008).

Gender gap according to disciplines

Teseo classifies dissertations according to the categories extracted from the UNESCO Nomenclature for fields of science and technology (UNESCO, 1988). This controlled vocabulary was originally intended to improve the exchange of statistical information on science but has also been used for the organization of scientific and technical literature. Although it has been used widely, it has not been updated since 1988 (Ruiz-Martínez et al., 2014), which has resulted in clear misadaptation to some scientific fields, such as Planetary Geology and Astrobiology (Martínez-Frías & Hochberg, 2007), Communication (Marzal Felici et al., 2016), Nursing (Pedraz Marcos, 2005) and even some medical disciplines (Fernández-Guerrero, 2015). The UNESCO Nomenclature was used to classify dissertations in Spain, but was also formally adopted and used as a reference in scientific policy-making, which made it worthy of attention to scholars over the years.

The UNESCO Nomenclature is divided into three levels (field, discipline, and sub-discipline) that refer to increasingly narrower levels of specialization. Indexing practices have varied over time, as noted by Sánchez-Jiménez et al. (2017), but in practice, in Teseo, two or more categories were assigned to most of the records (76%). As there are no consistent indexing practices, some theses were assigned several categories of different levels inside the same field. Those categories were often nested, as they would describe a scientific specialization from its broader to its narrower levels. Some others were more useful, identifying several specific scopes inside the same field, and some theses were assigned categories from different fields, which would imply a certain degree of multidisciplinary.

Detailing the structure of knowledge and the development of interdisciplinary approaches in doctoral dissertations is interesting but requires a greater level of detail in the analysis, and solving further problems related both to the assignment of categories, the structure of the controlled vocabulary and its relatedness to newer, more widely accepted scientific classifications. We opted for a simple procedure that classifies dissertations according to first level categories (fields). Fields were either extracted or inferred from the structure of the nomenclature. That is, a dissertation indexed under a specialization code

would be assigned the field code instead. This procedure destroys rich information but provides a more solid base for our next analysis.

In general, there has been a sharp increase in the number of females involved in the examination process over the last years. Figure 2 mixes data of all the roles that individuals had during years 2000 and 2020, as a way to examine the general patterns affecting disciplinary of the gender bias. In 12 out of the 24 fields the percentage of women doubled or was very close to doing it over the period. Of these fields, Medicine, Sociology and Juridical Sciences seem especially important. Both Medicine (16%) and Juridical sciences (18%) had a low presence of women by 2000, which has increased by more than 20%. Sociology was already in the high end of disciplines that had a greater female presence, but that presence has increased more than in any other field, apart from Logic.

The percentage of women involved in the areas of Mathematics and Physics grew too, but much less than the rest of the fields, apart from Geography and Ethics, which are on the other hand much less significant in the scientific landscape. Other disciplines that would be considered STEM are also increasing their percentage of participation of women, but at a lower rate than most of the rest of the fields. There is a huge literature on the gender gap regarding these areas, whether this refers to the early stages of education, jobs in science or the private sector (Ayuso et al., 2020; de las Cuevas et al., 2022; Makarova et al., 2019; McNally, 2020; Tandrayen-Ragoobur & Gokulsing, 2021). Therefore we devoted special attention to the STEM fields.

We are well aware that our definition of STEM disciplines is not universal, but as the delineation of STEM disciplines is subject to debate, we opted to choose fields that we found in the core of some important classifications. The Higher Education Statistics Agency of the UK defines STEM disciplines according to the Joint Academic Coding System (HESA, 2013). This includes Medicine (and other subjects allied to medicine), Biological Sciences, Veterinary, Agriculture, Physical Sciences, Mathematics, Computer Science, Engineering & Technology and Architecture. The NSF, on the other hand uses the term “Science and Engineering”, which is somehow coincident, but does not include medicine, veterinary or architecture, and it does include Social Sciences and Psychology

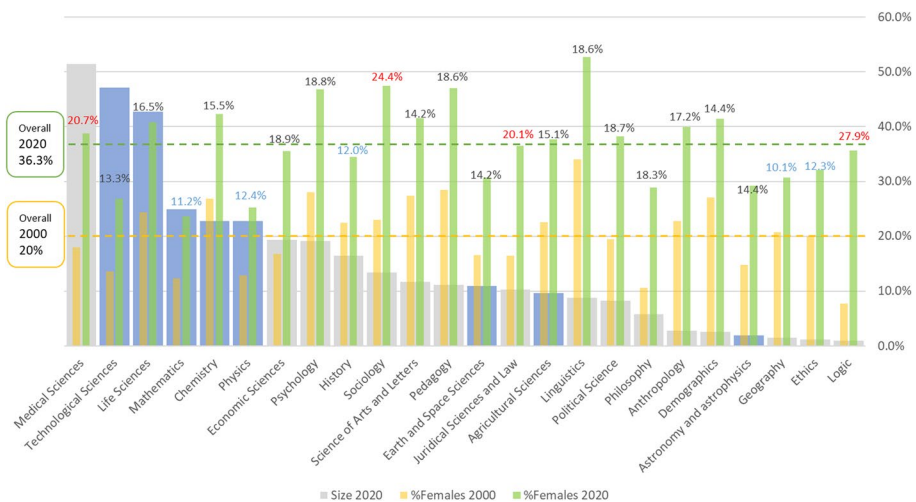


Fig. 2 Percentage of females in the doctoral system distributed by scientific fields, 2000 vs 2020

(NSF, 2021). We have opted for a definition of STEM that includes only the common fields, which implies that prominent fields as Medicine and Social Sciences are considered separately.

Two important changes can be observed in Fig. 2, STEM wise. If we consider the differential between percentages in 2000 and percentages in 2020 (the number overlaid on the respective columns) all the STEM fields (shaded in blue) are below the average increase of 16.6%. The gap between those fields and the overall percentage of women has increased over the last 20 years. That is, if the percentage of one of these fields was below the overall percentage, it has gone even lower. If it was higher than the overall percentage, it might still be so, but the difference has reduced over the years. Only Life Sciences has kept a similar distance to the overall percentage during the analyzed period.

This is a very general picture of the situation, and although it implies a significant advance in the gender balance of the highest level of education, and offers some interesting insights, it also hides some important information. The metaphor of the leaky pipeline has been applied many times to describe the decreasing number of females that reach the highest educational (Blackburn, 2017), professional (Ahuja, 1995) or scientific levels (Blickenstaff, 2005), or even achieve a successful transition to scientific careers (White, 2004). The study of this phenomenon has been linked to the STEM fields right from the beginning (Berryman, 1983), and although things might have changed over time, it is still actively studied.

If we apply this metaphor to doctoral education, we could consider the different roles as a pipeline that begins with the candidates, then continues with members of the committees (which might or not be junior academics but have usually made their way into the academy), then progresses towards positions of greater responsibility (supervisors) and arrive to positions of both responsibility and seniority (chairs of the committee). This pattern was already visible while analyzing the evolution of the different roles, but a clearer and more detailed picture can be obtained from its depiction in Fig. 3.

If we analyze the different roles of the individuals over the different scientific fields, we can see important differences between the two snapshots (again, 2000 vs 2020). The percentage of women has risen in all the roles, and the difference between the first and last sections of the pipeline has smoothed. This seems especially true for the ten biggest fields (according to the number of successful candidates). At the same time, a deeply rooted general pattern prevails. There is an increasingly lower number of women in increasing higher rank positions in the doctoral system.

This is not true, however, for every field. This staggered structure is much less clear for the Technological Sciences and Physics and is no longer recognizable in Astronomy and Astrophysics or Mathematics. The bad news is that the presence of women in these categories is among the lowest and seems to have stopped growing. Geography in the Social Sciences and Logic in the Humanities seem to not be showing this staggered pattern either, although this fields are much less significant, due to their reduced sizes.

In general, the imbalance between the lower role and the rest of the roles has decreased substantially in almost every field. However, the transition from members to supervisors has seen a growth in the gap between the percentage of females in each role. The difference between chairs and members has evolved differently according to the disciplines. In some cases, it has widened, whilst in others it has narrowed. The gap between the percentage of members and the percentage of chairs has not been reduced substantially during the last years. These two roles seem to have evolved pairwise during the period. The same goes to the percentage of candidates and the percentage of supervisors, which also evolve very similarly.

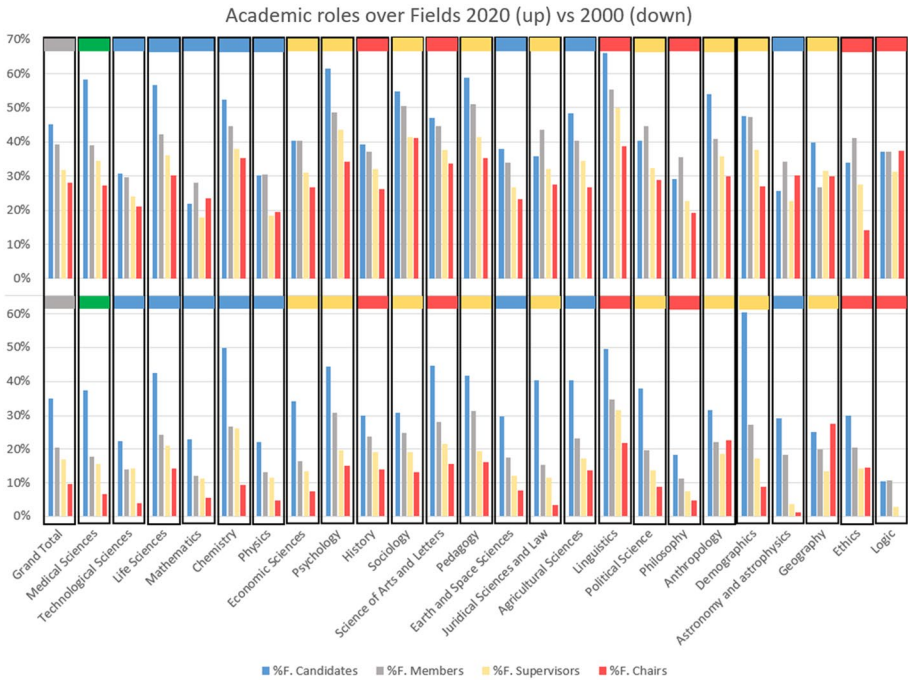


Fig. 3 Distribution of roles over fields, 2000 vs 2020

The committees are formed by an increasing number of women. As discussed before, there seems to be no significant increase in the percentage of female committee members after the new rules enter into force and Chairing in those committees seems to be performed by women according to a fraction of the available female academics. That is, it seems that either demography or a change in academic culture are playing a greater role than a negative gender bias in selecting members and chairs in the committees. At the same time, the percentage of women acting as supervisors seems to be flattening if not reducing, much as the percentage of candidates that elect them, both globally and at a field level, at least in most of the cases. Figure 4 offers a general view of this evolution over the four main branches of science. This led us to think that there might be a gender assortativity pattern pushing the number of female supervisors for the last years, which is discussed below.

Another worrisome pattern that emerges from Fig. 4 is the widening of the gap between female members and female supervisors. This would be a consequence of the other two patterns that we have highlighted but creates a counter-intuitive and regressive trend for the future. This gap is around 10% for all the branches according to the last data, but while in Medical sciences this trend seems recent and might not consolidate, in STEM seems to have been going on for a longer period. In the rest of the cases, the gap has existed for the last 30 years, and has not varied that much until recently.

Another visible difference between the branches is the gap between the percentage of female candidates and the percentage of female members. This has reduced extraordinarily overtime and has but disappeared, except for the case of the Medical Sciences, in which this is still very substantial (17%). The gap between the percentage of female supervisors

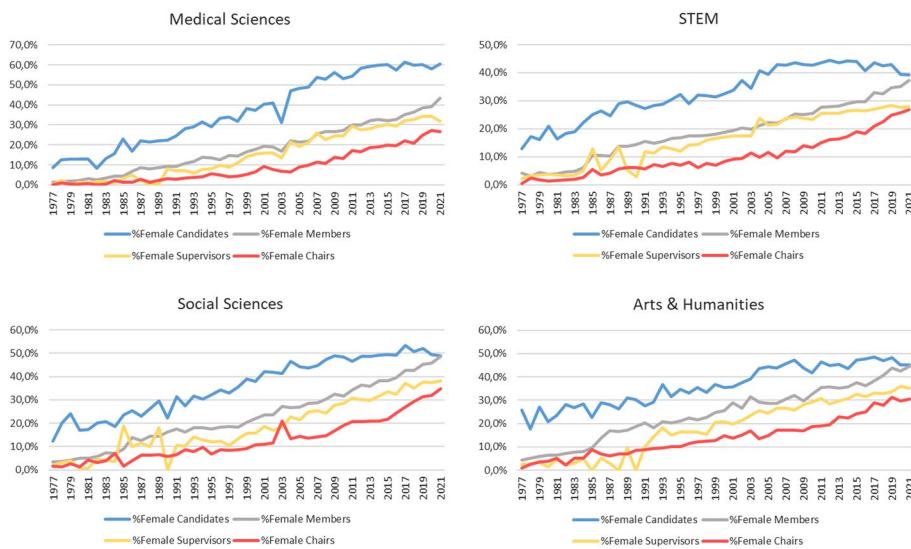


Fig. 4 Evolution of doctoral roles in Medical Sciences, STEM, Social Sciences and Humanities

and the percentage of female chairs is around 5% or lower, and has basically disappeared in the case of STEM.

Institutional distribution

Spanish universities offer a certain degree of diversity in that the system is formed by both private (41) and public institutions (50). Some of these are technical or humanistic/social science driven institutions, although most of them try to encompass the global aspects of science, both from a teaching and research perspective. Some of the private universities (16) have no doctoral production yet and are not included in the representation below (Fig. 5).

We analyzed the proportion of candidates/supervisors in the Spanish universities for the whole period and distinguished between public/private and technical/generalist institutions (Table 3). As a whole, there is not a significant difference between private and public institutions regarding the percentages of female candidates or female supervisors. Technical universities had visibly lower participation of female candidates, whether private (28.9%) or public (27.8%), while generalist universities had a very similar average percentage of female candidates, which was significantly higher than technical institutions (41% vs 43%). It’s worth noting, though, that there is a single example of technical private university, the MU (Mondragon Unibertsitatea). The same pattern was found in the percentages of female supervisors. Technical universities had a mere 18% (private) or 17% (public) of female supervisors, while generalist universities showed significantly higher percentages, both in the case of the private institutions (26%) and the public institutions (25%).

It is worth noting that the PHD programs of private universities account only for a 5.5% of the candidates included in the study. Of these, the Universidad de Navarra (UN in Fig. 5) accounts for roughly half of the successful candidates (48%). This university in particular shows a quite distinctive pattern, with simultaneously low percentages of female

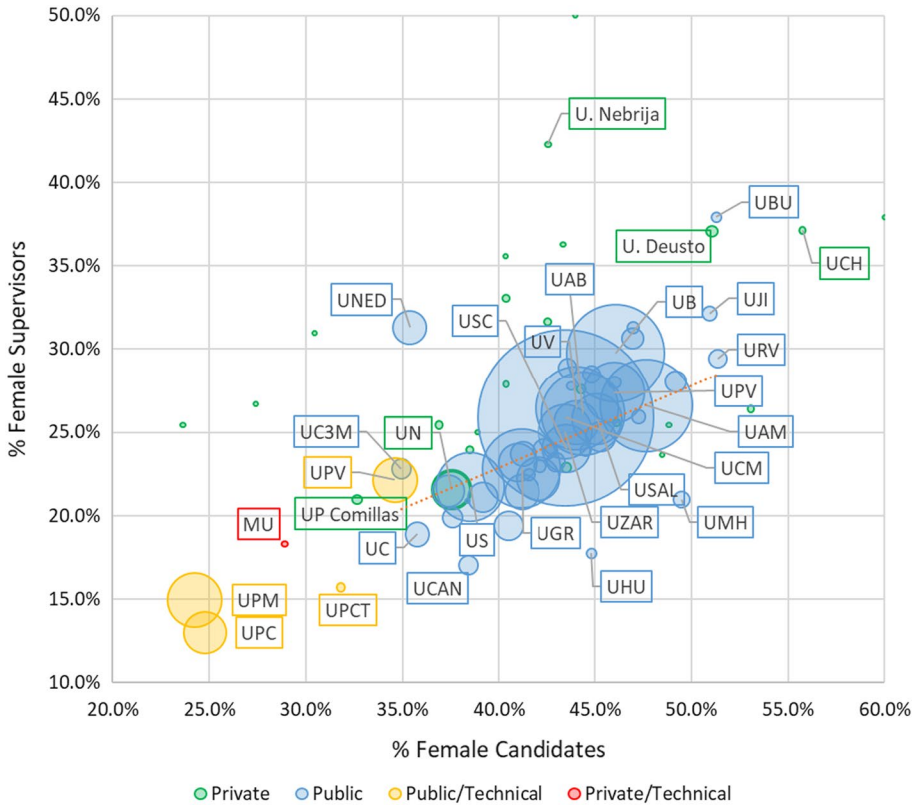


Fig. 5 Accumulated percentages of female candidates vs female supervisors in Spanish universities

Table 3 Average percentages of female academics per role and type of institution

Type of institution	%F. Candidates	%F. Members	%F. Supervisors	%F. Chairs
Private	41	25.6	26.4%	14.5%
Public	43.3	25.8	25.2%	15.3%
PrivateTech	28.9	19.7	18.3%	15.1%
PublicTech	27.8	17.1	16.8%	10.6%

supervisors (21%) and female candidates (37%). The rest of the private institutions are much smaller which probably enables the high variability of female presence in both candidate and supervisor roles.

Figure 5 shows a much more solid pattern for the public universities, while private ones seem much more variable, as we anticipated. Technical universities can be found (however) in the lower left quadrant without exceptions, corroborating the general impression that aggregated figures gave us, which is also consistent with what we know about the distribution of female candidates and supervisors among STEM fields. Its worth mentioning that although the patterns are clear, important differences exist between institutions in the same group. UB and US, for instance, are visibly apart in terms of both female candidates and

supervisors. The UNED shows is also worth mentioning, as is quite apart from the rest of the universities, with an almost unexisting gap between the percentages of supervisors and candidates. The case of the Universidad Politécnica de Valencia (UPV) is also relevant, as it shows a clear distance from the rest of the technical universities, although it's located at the end of the queue of generalist universities. This would imply that local policies or institutional idiosyncrasies could play a role in shaping gender imbalance around the doctoral system.

Gender assortativity

Homophily or gender assortativity has been detected both in theoretically symmetric relationships, such as coauthorship, and in asymmetric relationships, such as the supervisor/supervisee relationship. Villarroya et al (2008) confirmed that this phenomenon could be found in the Spanish doctoral system, but their study described a time in which the presence of women in both examination committees and supervisorship was much more reduced and co-supervision was less common. We thought that the greatly increased availability of female academics, which were already participating in great numbers in the committees, could make an update interesting.

As of 2021, the percentage of female candidates that were supervised by a female supervisor (23%) was still very far from the percentage of female candidates that had a male supervisor (45%). Even though, the situation has evolved significantly, and we can see an important change from the beginning of the century. In 2001, the proportion would have been 16% to 71%. Similarly, male candidates are supervised by male supervisors much more frequently (63%) than they are supervised by female counterparts (13%). Again, this is a significant change from 2001, when percentages were 83% to 9%.

For the same year (and also the same number of male/female supervisors) female candidates have been proportionally more associated with female supervisors (23%) than male candidates have been (13%). This pattern has been going on for a long time and seems to point to the existence of a preference for same gender associations between candidates and supervisors.

Another interesting pattern that we can see in Fig. 6 is the growth of mixed gender supervising. Mixed supervision seems to be a phenomenon of the 2000's, as it was basically anecdotal before. This kind of supervision is more common among female candidates (31% of the theses in 2021), although is also significant in the case of male candidates (24%). It is difficult anyhow to make out whether these figures would be above the expected percentages, so we tested the idea using a contingency table.

Table 4 offers an overview of the significance of the association between the gender of the candidates and the gender of the supervisors. Both female and male candidates are more prone to be associated with supervisors of the same gender. In the case of female candidates, they are also significantly more prone to be associated with mixed teams of supervisors than the male candidates. This confirms what we already knew for a more recent period of time (2017–2021), but also informs of an increase in the level of association as measured with Cramer's V, from 0,14 (Villarroya et al., 2008) to 0,17. This would mean that homophily has grown over time.

Another issue that seems worthy of attention is the relation between the gender of the supervisors and the gender of the committee chairs. Although the role of supervisors is clearly much more important in shaping the future of candidates, the committees do also play a role in introducing candidates to a network of potential collaborators (Bès et al.,

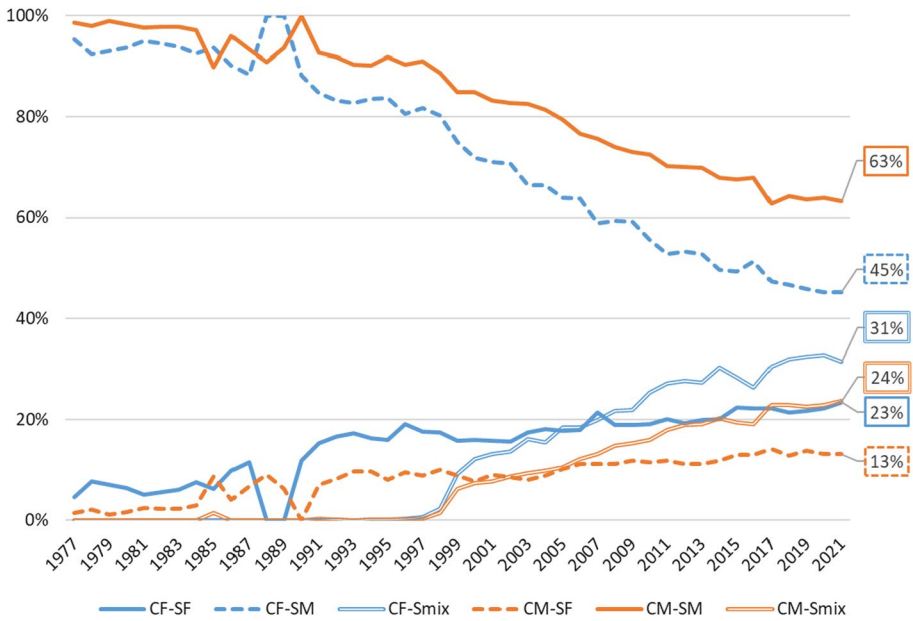


Fig. 6 Percentage of theses according to the gender of the supervisor (SF/SM/Smix) regarding the total number of theses with female candidates (CF, in blue) and the total number of theses with male candidates (CM, orange)

Table 4 Relationship between the gender of the candidates and the gender of the supervisors (2017–2021)

2017–2021	Gender of the SUPERVISOR			Total	X ² (d.f)	p	C's V
	Female	Male	Mixed				
Female Candidate	%	10,7%	22,5%	15,3%	49%	1055(2)	< 0.000001
	N	3740	7854	5343			
	SR	13,465	-15,414	11,102			
Male Candidate	%	7,0%	32,6%	11,8%	51%		
	N	2436	11,368	4101			
	SR	-13,096	14,992	-10,797			
Total	%	17,7%	55,2%	27,1%	100%		
	N	6176	19,222	9444			

N: number of theses, SR: Standardized residuals, X2: chi-square test, d.f: degrees of freedom, p: level of significance, C'V: Cramer's V coefficient.

2021) and can have a relevant paper in introducing new ideas, as the process of examination is also a “vehicle for scientific exchange” (Breimer, 2013). In this process, the chairs of the committee have a relevant role, and they are also supposed to be prominent from a scientific point of view, as well as regarding the underlying social structures that can be detected analyzing committee networks (Olmeda et al., 2009). The reported 1% of female chairs (Villarroya et al. AI, 2008) has changed very substantially overtime, so we thought that the relation between the gender of the supervisor and the gender of the committee chair might be interesting to study (Fig. 7).

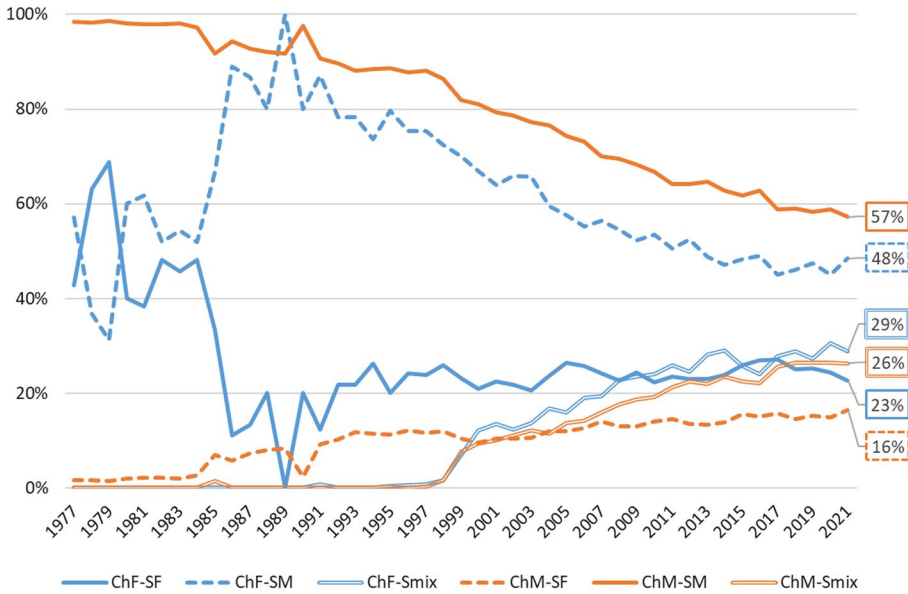


Fig. 7 Percentage of these according to the gender of the supervisor (SF/SM/Smix) regarding the total number of these with female chairs (ChF, in blue) and the total number of these with male chairs (ChM, orange)

As we can see, the situation has changed from the early 2000’s on, but not exactly as the relationship between candidates and supervisors. As of 2021, the percentage of female chairs that were associated with female supervisors (23%) was still very far from the percentage of female chairs that were associated with male supervisors (48%). If we compare these percentages to the ones of the first 2000’s when the proportion would have been 22% to 64% (2001), we can see that only the percentage of male supervisors/female chairs has reduced significantly. The factor that would explain this would be mixed-gender teams of supervisors, that for 2021 account for a 29% of the committees chaired by females.

Regarding the idea of a gender preference, female supervisors coincide with female chairs much more (23%) than with male chairs (16%), which would indicate again homophily between chairs and supervisors. We have also created a contingency table to deepen into this idea (see Table 5).

Table 5 shows that there is a significant association between the gender of the committee chairs and the gender of the supervisors. Both female and male chairs are more prone to be associated with supervisors of the same gender. In the case of female chairs, they are also more prone to be associated with mixed-gender teams of supervisors than the male chairs.

Possible limitations

The matching of mentions of individuals and persons’ records is still rudimentary, and homonym academics could not be resolved at this stage. We think that their weight regarding the rest of individuals would be very limited, though. Name identification and gender

Table 5 Relationship between the gender of the Committee Chairs and the gender of the supervisors (2017–2021)

2017–2021		Gender of the supervisor			Total	X ² (d.f)	p	C's V
		Female	Male	Mixed				
Female Chair	%	6,6%	11,8%	7,3%	26%	623(2)	< 0.000001	0,12,981
	N	2423	4375	2705	9503			
	SR	17,382	– 12,227	3,3587				
Male Chair	%	11,4%	43,5%	19,4%	74%			
	N	4212	16,099	7162	27,473			
	SR	– 10,223	7,1909	– 1,9754				
Total	%	17,9%	55,4%	26,7%	100%			
	N	6635	20,474	9867	36,976			

N number of theses, *SR* Standardized residuals, *X*² chi-square test, *d.f.* degrees of freedom, *p* level of significance, *C'V* Cramer's V coefficient.

assignment are also clearly improvable, and although the distribution of gender of the individuals that could not be assigned one does not appear to be clearly biased, further work is needed to root this out or correct it. Also, Teseo has limitations of its own, that have been highlighted in the data and method's section. These limitations are well described in the literature and could lead to localized imprecisions, although the database is widely regarded as a relevant source of information on the Spanish doctoral system.

Discussion and conclusions

The gap between males and females in the doctoral system has narrowed very much over the years, although important challenges and worrying trends can still be detected when considering data from some angles.

The gender of successful candidates seems to have reached balance during the last years, but the percentage of female candidates has also decreased to 2010 levels after that. It's still soon to decide whether this is an important pattern, but it should probably be tracked in the future. The percentage of female supervisors it's also much higher than it was some years ago, but its growth has stalled, as it has remained at very similar levels for the last five years. This pattern is clearer in some disciplines, while in others it's harder to appreciate or involves a more limited time span. The case of STEM fields shows a more worrying pattern in this regard.

Female committee members and chair members, on the other hand, have clear upwards trends during the whole period. The committees are formed by an increasing number of women, and chairing in those committees seems to be performed by women according to a fraction of the available female academics. Governmental action towards parity in the committee's system does not seem to have had any effect on the percentage of women that were chosen to participate as members or chairs of the committees. We hypothesize that the reduction of gender imbalance in both cases might be related to organic factors such as the growth of the number of female academics, but also a probable change in academic culture.

The percentage of female academics is different according to the different roles, and this difference seems to have built a structure of echelons from the lower roles (in terms of seniority and responsibility in the doctoral process) to the higher ones. Women constitute a great percentage of the candidates, but are progressively less present in the following roles, as members of the committee or supervisors, and are clearly a minority in committee chairing roles. This structure seems to fit in the *leaky pipeline* metaphor, as we discussed above. It also seems less prominent in the last years than it used to be but is still clearly visible in all the scientific branches and has only disappeared in some of the fields. This could be caused by different academic cultures in the respective disciplines, but other factors could also explain this.

The presence of female academics (in any of the roles) has seen a very important increase over the last years in all of the 24 scientific fields in which theses were classified. This trend has been going on at different paces, nevertheless. It's worth noting that gender imbalance has been reduced less in case of the STEM branch than in other branches. STEM fields have been reducing gender imbalance at a lower rate than most of the rest of the fields.

Regarding the institutional distribution of gender imbalance, the type of university seems to be an important factor in explaining it. Technical universities show a lower percentage of females in any role, while universities with a global scope tend to have greater gender balance. The difference between private and public institutions does not seem significant, though. In both cases there is an important variability between institutions, which gives rise to our hypothesis that local efforts towards gender balance might have a relevant and measurable impact in avoiding a gender bias in the doctoral system.

Mixed-gender supervising teams is a phenomenon that was also described in the past and began to have relevance at the beginning of the 2000's but has evolved significantly in the last times. This kind of setting is today much more frequent, so much so that candidates that do not have a supervisor of the same gender tend to have supervisor teams of mixed gender much more often than single supervisors of a different gender. The increasing importance of collaboration among supervisors is obviously a factor, but the fact that mixed gender supervisor teams are much more common than female supervisors acting alone is worth studying.

Gender assortativity was clearly found in our data, both in the case of the relations between supervisors and candidates and among supervisors and committee chairs. Female candidates tend to be associated with female supervisors much more than their male counterparts are associated with female supervisors. Female committee chairs are much more common when the supervisor is a female than in the case when the supervisor is male. Again, mixed gender supervisor teams are more often associated with female chairs than they are with their male counterparts. The meaning of this is not clear to us, and hypothesizing would require further data, probably through a qualitative approach, but it seems like a clear pattern that would be worth studying. In both the chair/supervisor and candidate/supervisor relationships the assortativity has grown over the years, and might be worth of a follow-up in the future.

Overall, the most concerning issues about gender balance in the Spanish doctoral system seem to be linked to STEM fields and to the growth of the female involvement in some of the roles. This is not surprising, as the study of these disciplines has been the focus of great attention in the literature, given the specific difficulties of diminishing the gender gap in the domain. Homophily, which was detected in the past, seems to have grown significantly, and could be having important effects in several evolving patterns,

such as the developments around the leaky pipeline or the dynamics around supervising teams. As such, this is probably the most interesting line of research for the future.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest The authors declare they have no conflicts of interest to report.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009). Gender differences in research productivity: A bibliometric analysis of the Italian academic system. *Scientometrics*, 79(3), 517–539. <https://doi.org/10.1007/s11192-007-2046-8>
- Ahuja, M. K. (1995). Information technology and the gender factor. *Proceedings of the 1995 ACM SIGCPR conference on Supporting teams, groups, and learning inside and outside the IS function reinventing IS*, 156–166. <https://doi.org/10.1145/212490.212595>
- Araújo, T., & Fontainha, E. (2017). The specific shapes of gender imbalance in scientific authorships: A network approach. *Journal of Informetrics*, 11(1), 88–102. <https://doi.org/10.1016/j.joi.2016.11.002>
- Ayuso, N., Fillola, E., Masia, B., Murillo, A. C., Trillo-Lado, R., Baldassarri, S., Cerezo, E., Ruberte, L., Mariscal, M. D., & Villarroya-Gaudó, M. (2020). Gender gap in STEM: A cross-sectional study of primary school students' self-perception and test anxiety in mathematics. *IEEE Transactions on Education*, 64(1), 40–49. <https://doi.org/10.1109/TE.2020.3004075>
- Berryman, S. E. (1983). *Who will do science? Minority and female attainment of science and mathematics degrees: Trends and causes*. Rockefeller Foundation.
- Bès, M.-P., Lamy, J., & Maisonnobe, M. (2021). Peer-making: The interconnections between PhD thesis committee membership and copublishing. *Quantitative Science Studies*, 2(3), 1048–1070. https://doi.org/10.1162/qss_a_00143
- Blackburn, H. (2017). The status of women in STEM in higher education: A review of the literature 2007–2017. *Science & Technology Libraries*, 36(3), 235–273. <https://doi.org/10.1080/0194262X.2017.1371658>
- Blickenstaff, J. C. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education*, 17(4), 369–386. <https://doi.org/10.1080/09540250500145072>
- Bordons, M., Morillo, F., Fernández, M. T., & Gómez, I. (2003). One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists. *Scientometrics*, 57(2), 159–173. <https://doi.org/10.1023/A:1024181400646>
- Borrego, Á., Barrios, M., Villarroya, A., & Ollé, C. (2010). Scientific output and impact of postdoctoral scientists: A gender perspective. *Scientometrics*, 83(1), 93–101. <https://doi.org/10.1007/s11192-009-0025-y>
- Breimer, L. H. (2013). Impact of foreign external PhD examiners on academic collaboration patterns. *Scientometrics*, 96(1), 315–322. <https://doi.org/10.1007/s11192-012-0931-2>
- Breimer, L. H., & Mikhailidis, D. P. (2020). Half a century and more of PhD theses by published papers: Comment on: “Bringing the doctoral thesis by published papers to the Social Sciences and the Humanities: A quantitative easing? A small study of doctoral thesis submission rules and practice in two disciplines in the UK” by John Rigby and Barbara Jones in *Scientometrics* published online 15-May-2020. *Scientometrics*, 125(1), 813–816. <https://doi.org/10.1007/s11192-020-03622-2>

- Bu, Y., Li, H., Wei, C., Liu, M., & Li, J. (2020). On the relationship between supervisor–supervisee gender difference and scientific impact of doctoral dissertations: Evidence from Humanities and Social Sciences in China. *Journal of Information Science*, 48(4), 492–502. <https://doi.org/10.1177/0165551520969935>
- Castellano, J., Alvarez-Pastor, D., & Bradley, P. S. (2014). Evaluation of research using computerised tracking systems (amisco® and prozone®) to analyse physical performance in elite soccer: A systematic review. *Sports Medicine*, 44(5), 701–712. <https://doi.org/10.1007/s40279-014-0144-3>
- Castelló-Cogollos, L., Aleixandre-Benavent, R., & Castelló-Cogollos, R. (2015). Differences by gender and role in PhD theses on sociology in Spain. *Proceedings of ISSI*, 608–609. <https://doi.org/10.13039/501100003359>
- Catalá-López, F., Peiró, S., Ridao, M., Sanfèlix-Gimeno, G., Gènova-Maleras, R., & Catalá, M. A. (2012). Prevalence of attention deficit hyperactivity disorder among children and adolescents in Spain: A systematic review and meta-analysis of epidemiological studies. *BMC Psychiatry*, 12, 168. <https://doi.org/10.1186/1471-244X-12-168>
- Crane, Diana (1972). *Invisible colleges: diffusion of knowledge in scientific communities*. University of Chicago. ISBN: 0226118576
- Curriel-Marín, E., & Fernández-Cano, A. (2015). Scientometric Analysis of Spanish Doctoral Theses on the Teaching of Social Sciences (1976–2012). *Revista Española de Documentación Científica*, 38(4), e110. <https://doi.org/10.3989/redc.2015.4.1282>
- De las Cuevas, P., García-Arenas, M., & Rico, N. (2022). Why Not STEM? A Study Case on the Influence of Gender Factors on Students' Higher Education Choice. *Mathematics*, 10(2), 239. <https://doi.org/10.3390/math10020239>
- Dehdarirad, T., & Nasini, S. (2017). Research impact in co-authorship networks: A two-mode analysis. *Journal of Informetrics*, 11(2), 371–388. <https://doi.org/10.1016/j.joi.2017.02.002>
- Delgado López-Cózar, E., Torres-Salinas, D., Jiménez-Contreras, E., & Ruiz-Pérez, R. (2006). Análisis bibliométrico y de redes sociales aplicado a las tesis bibliométricas defendidas en España (1976–2002): Temas, escuelas científicas y redes académicas. *Revista Española De Documentación Científica*, 29(4), 493–524. <https://doi.org/10.3989/redc.2006.v29.i4.306>
- Diaz-Kope, L. M., Miller-Stevens, K., & Henley, T. J. (2019). An examination of dissertation research: The relationship between gender, methodological approach, and research design. *Journal of Public Affairs Education*, 25(1), 93–114. <https://doi.org/10.1080/15236803.2018.1463792>
- Etzkowitz, H., Kemelgor, C., Neuschatz, M., & Uzzi, B. (1992). Athena unbound: Barriers to women in academic science and engineering. *Science and Public Policy*, 19(3), 157–179.
- Fernández-Cano, A., Torralbo, M., & Vallejo, M. (2012). Time series of scientific growth in Spanish doctoral theses (1848–2009). *Scientometrics*, 91(1), 15–36.
- Fernández-Guerrero, I. M. (2015). Spanish doctoral theses in emergency medicine (1978–2013). *Emergencias*, 27(2), 129–134. <https://pubmed.ncbi.nlm.nih.gov/29077355/>
- Fernández-Guerrero, I. M., Callejas, Z., Griol, D., & Fernández-Cano, A. (2020). Longitudinal patterns in Spanish doctoral theses on scientific medical information: A tertiary study. *Scientometrics*, 124(2), 1241–1260. <https://doi.org/10.1007/s11192-020-03494-6>
- Fuentes Pujol, E., Pujol, E. F., & Vivó, L. A. (2010). Las tesis doctorales en España (1997–2008): Análisis, estadísticas y repositorios cooperativos. *Revista Española De Documentación Científica*, 33(1), 63–89. <https://doi.org/10.3989/redc.2010.1.711>
- Gaule, P., & Piacentini, M. (2018). An advisor like me? Advisor gender and post-graduate careers in science. *Research Policy*, 47(4), 805–813. <https://doi.org/10.1016/j.respol.2018.02.011>
- Green, C., Jegadeesh, N., & Tang, Y. (2009). Gender and job performance: Evidence from Wall Street. *Financial Analysts Journal*, 65(6), 65–78. <https://doi.org/10.2469/faj.v65.n6.1>
- González-Alcaide, Gregorio; González-Teruel, Aurora (2020). Dirección de tesis doctorales como reflejo de la Bibliotecología y Documentación académica española: actores y temas. *Profesional de la Información*, 29(4), e290441. <https://doi.org/10.3145/epi.2020.jul.41>
- Hernández-González, V., De Pano-Rodríguez, A., & Reverter-Masia, J. (2020). Spanish doctoral theses in physical activity and sports sciences and authors' scientific publications (LUSTRUM 2013–2017). *Scientometrics*, 122(1), 661–679. <https://doi.org/10.1007/s11192-019-03295-6>
- Hernández-Martín, E., Calle, F., Dueñas, J. C., Holgado, M., & Gómez-Pérez, A. (2019). Participation of women in doctorate, research, innovation, and management activities at Universidad Politécnica de Madrid: Analysis of the decade 2006–2016. *Scientometrics*, 120(3), 1059–1089. <https://doi.org/10.1007/s11192-019-03179-9>
- HESA. (2013). *Joint Academic Coding System: Version 3.0*. <https://www.hesa.ac.uk/support/documentation/jacs>

- Holman, L., & Morandin, C. (2018). Researchers collaborate with same-gendered colleagues more often than expected across the life sciences (p. 345975). *bioRxiv*. <https://doi.org/10.1101/345975>
- Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLoS Biology*, 16(4): e2004956. <https://doi.org/10.1371/journal.pbio.2004956>
- Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, 117(9), 4609–4616. <https://doi.org/10.1073/pnas.1914221117>
- INE. (2021b). *Estadística de Tesis Doctorales (ETD). Año 2020*. <http://estadisticas.mecd.gob.es/EducaDynPx/educabase/index.htm?type=pcaxis&path=/Universitaria/Tesis/2020/&file=pcaxis>
- INE. (2021a). *Apellidos y nombres más frecuentes*. Latest data from year 2020. Published: 20/05/2021b and 17/06/2021b. <https://www.ine.es/uc/ijPGiEWy>
- INE. (2021c). *Mujeres en el profesorado por enseñanza que imparten*. Latest data from the 2018/2019 academic year. <https://www.ine.es/uc/bNIGk16f>
- Jadidi, M., Karimi, F., Lietz, H., & Wagner, C. (2018). Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. *Advances in Complex Systems*, 21(3/4), 1750011. <https://doi.org/10.1142/S0219525917500114>
- Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M., & Strohmaier, M. (2016). Inferring gender from names on the web: A comparative evaluation of gender detection methods. *Proceedings of the 25th International conference companion on World Wide Web*, 53–54. <https://doi.org/10.1145/2872518.2889385>
- Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Global gender disparities in science. *Nature*, 504, 211–213. <https://doi.org/10.1038/504211a>
- Government of Spain (2007). Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres. *Boletín Oficial del Estado*, 71, de 23 de marzo de 2007. <https://www.boe.es/buscar/act.php?id=BOE-A-2007-6115>
- López-Yepes, José (2002). “Focos de investigación y escuelas científicas en documentación. La experiencia de las tesis doctorales”. *El Profesional de la Información*, 11(1), 46–52. <http://www.elprofesionalde lainformacion.com/contenidos/2002/enero/6.pdf>
- Macaluso, B., Larivière, V., Sugimoto, T., & Sugimoto, C. R. (2016). Is science built on the shoulders of women? A study of gender differences in contributorship. *Academic Medicine*, 91(8), 1136–1142. <https://doi.org/10.1097/ACM.0000000000001261>
- Makarova, E., Aeschlimann, B., & Herzog, W. (2019). The gender gap in STEM fields: The impact of the gender stereotype of math and science on secondary students’ career aspirations. *Frontiers in Education*, 4, 60. <https://doi.org/10.3389/educ.2019.00060>
- Martínez-Frías, J., & Hochberg, D. (2007). Classifying science and technology: Two problems with the UNESCO system. *Interdisciplinary Science Reviews*, 32(4), 315–319. <https://doi.org/10.1179/030801807X183605>
- Marzal Felici, J., García Jiménez, A., & Humanes, M. L. (2016). Análisis y reformulación de la organización del conocimiento en las Ciencias de la Comunicación: Aplicación para la codificación UNESCO. *Revista General De Información y Documentación*, 26(1), 65–79. https://doi.org/10.5209/rev_RGID.2016.v26.n1.53049
- McNally, S. (2020). *Gender differences in tertiary education: What explains STEM participation?* IZA Policy Paper, No. 165. Institute of Labor Economics (IZA). <https://www.econstor.eu/handle/10419/243451>
- Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. (2011). Understanding the Demographics of Twitter Users. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 554–557. <https://doi.org/10.1609/icwsm.v5i1.14168>
- Musi-Lechuga, B., Olivás-Ávila, J. A., & Buéla-Casal, G. (2009). Producción científica de los programas de Doctorado en Psicología Clínica y de la Salud de España. *International Journal of Clinical and Health Psychology*, 9(1), 161–173. <http://hdl.handle.net/10481/33092>
- NSF. (2021). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2021*. National Science Foundation. <https://ncses.nsf.gov/pubs/nsf21321/report/field-of-degree-women>
- Olmeda-Gómez, C., Perianes-Rodríguez, A., Ovalle-Perandones, M. A., & De-Moya-Anegón, F. (2009). Colegios visibles: Estructuras de coparticipación en tribunales de tesis doctorales de biblioteconomía y documentación en España. *El Profesional De La Información*, 18(1), 41–49. <https://doi.org/10.3145/epi.2009.ene.06>
- Pedraz Marcos, A. (2005). Enfermería no está en el Código UNESCO. *NURE Investigación*, 13. <https://www.nureinvestigacion.es/OJS/index.php/nure/article/view/218>
- Repiso-Caballero, R., Torres-Salinas, D., & Delgado López-Cózar, E. (2012). Análisis de la investigación sobre Radio en España: Una aproximación a través del Análisis Bibliométrico y de Redes Sociales de

- las tesis doctorales defendidas en España entre 1976–2008. *Estudios Sobre El Mensaje Periodístico*, 17(2), 417–429. https://doi.org/10.5209/rev_ESMP.2011.v17.n2.38123
- Repiso-Caballero, R., Torres-Salinas, D., & Delgado López-Cózar, E. (2011). Bibliometric and social network analysis applied to television dissertations presented in Spain (1976/2007). *Comunicar*, 37(9), 151–159. <https://doi.org/10.3916/C37-2011-03-07>
- Reybold, L. E., Brazer, S. D., Schrum, L., & Corda, K. W. (2012). The politics of dissertation advising: How early career women faculty negotiate access and participation. *Innovative Higher Education*, 37(3), 227–242. <https://doi.org/10.1007/s10755-011-9200-1>
- Ruiz-Martínez, J.-M., Baños-Moreno, M.-J., & Martínez-Béjar, R. (2014). Nomenclatura Unesco: Evolución, alcance y reutilización en clave ontológica para la descripción de perfiles científicos. *El Profesional De La Información*, 23(4), 383–392. <https://doi.org/10.3145/epi.2014.jul.06>
- Sánchez Jiménez, R., Blázquez Ochando, M., Montesi, M., & Botezan, I. (2017). La producción de tesis doctorales en España (1995–2014): evolución, disciplinas, principales actores y comparación con la producción científica en WoS y Scopus. *Revista Española de Documentación Científica*, 40(4): e188. <https://doi.org/10.3989/redc.2017.4.1409>
- Schiebinger, L. (1987). The history and philosophy of women in science: A review essay. *Signs: Journal of Women in Culture and Society*, 12(2), 305–332
- Schluter, N. (2018). The glass ceiling in NLP. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2793–2798. <https://doi.org/10.18653/v1/D18-1301>
- Tandrayen-Ragoobur, V., & Gokulsing, D. (2021). Gender gap in STEM education and career choices: What matters? *Journal of Applied Research in Higher Education*, 14(3), 1021–1040. <https://doi.org/10.1108/JARHE-09-2019-0235>
- Tomassini, C. (2021). Gender gaps in science: systematic review of the main explanations and research agenda. *Education in the Knowledge Society (EKS)*, 22: e25437. <https://doi.org/10.14201/eks.25437>
- UNESCO. (1988). *Proposed international standard nomenclature for fields of science and technology*. UNESCO/NS/ROU/257 rev.1. <https://unesdoc.unesco.org/ark:/48223/pf0000082946>
- Vallejo, M., Torralbo, M., & Fernández-Cano, A. (2016). Gender bias in higher education: Spanish doctoral dissertations in mathematics education. *Journal of Hispanic Higher Education*, 15(3), 205–220. <https://doi.org/10.1177/1538192715592927>
- Villarroya, A., Barrios, M., Borrego, A., & Frías, A. (2008). PhD theses in Spain: A gender study covering the years 1990–2004. *Scientometrics*, 77(3), 469–483. <https://doi.org/10.1007/s11192-007-1965-8>
- Wang, C., Guo, F., & Wu, Q. (2021). The influence of academic advisors on academic network of Physics doctoral students: empirical evidence based on scientometrics analysis. *Scientometrics*, 126(6), 4899–4925. <https://doi.org/10.1007/s11192-021-03974-3>
- Welsh, E., & Abramson, A. (2018). A measure of intra-university collaboration: Faculty gender imbalance on doctoral dissertation committees in engineering disciplines. *International Journal of Doctoral Studies*, 13, 457–469. <https://doi.org/10.28945/4141>
- West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The Role of Gender in Scholarly Authorship. *PLOS ONE*, 8(7), e66212. <https://doi.org/10.1371/journal.pone.0066212>
- White, K. (2004). The leaking pipeline: Women postgraduate and early career researchers in Australia. *Tertiary Education and Management*, 10(3), 227–241. <https://doi.org/10.1080/13583883.2004.9967129>
- Ying, T., & Xiao, H. (2012). Knowledge linkage: A social network analysis of tourism dissertation subjects. *Journal of Hospitality & Tourism Research*, 36(4), 450–477. <https://doi.org/10.1177/1096348011400745>
- Zong, Q.-J., Shen, H.-Z., Yuan, Q.-J., Hu, X.-W., Hou, Z.-P., & Deng, S.-G. (2013). Doctoral dissertations of Library and Information Science in China: A co-word analysis. *Scientometrics*, 94(2), 781–799. <https://doi.org/10.1007/s11192-012-0799-1>
- Zuckerman, H., & Cole, J. R. (1975). Women in American Science. *Minerva*, 13(1), 82–102.