Research Policy 51 (2022) 104561

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

What makes a productive Ph.D. student?

Alberto Corsini [a,*], Michele Pezzoni [a,b,c,d], Fabiana Visentin [e]

[a] Université Côte d'Azur, CNRS, GREDEG, France
[b] Observatoire des Sciences et Techniques, HCERES, France
[c] OFCE, SciencesPo, France
[d] ICRIOS, Bocconi University, Italy
[e] UNU-MERIT, Maastricht University, Maastricht, the Netherlands

ARTICLE INFO

JEL classification:
J24
O30

Keywords:
Ph.D. students
Productivity determinants
Social environment
Supervisor
Peers

ABSTRACT

This paper investigates how the social environment to which a Ph.D. student is exposed during her training relates to her scientific productivity. We investigate how supervisor and peers' characteristics are associated with the student's publication quantity, quality, and co-authorship network size. Unique to our study, we cover the entire Ph.D. student population of a European country for all the STEM fields analyzing 77,143 students who graduated in France between 2000 and 2014. We find that having a productive, mid-career, low-experienced, female supervisor who benefits from a national grant is positively associated with the student's productivity. Furthermore, we find that having few productive freshman peers and at least one female peer is positively associated with the student's productivity. Interestingly, we find heterogeneity in our results when breaking down the student population by field of research.

> "My supervisor has everything I was looking for in a mentor. She is young and ambitious, and she overcomes any inexperience with a thirst for sharing her knowledge. Choosing me as her first PhD while establishing her own research group, filled me with a sense of responsibility while giving me the freedom to create something that I consider my own."[1]
>
> (Testimonial by a second-year Ph.D. in Human Medicine)

> "Professor A's group has developed many multidisciplinary research frontiers. From his connections, I have the opportunities to work with excellent colleagues in the School of Medicine. The collaborative research experiences during my PhD study are beneficial for me to expand my expertise toolkit. All the group members in Professor A's lab are very productive and the atmosphere in the group has been very enjoyable. The size of the group is just right, and the group is very dynamic and collaborative."[2]
>
> (Testimonial by a graduate student in Electrical engineering)

1. Introduction

In the last 20 years, the OECD countries almost doubled the number of graduate students, passing from 154,000 in 2000 to 276,800 in 2017 (OECD, 2013, 2019), while the number of high skills job positions did not increase at the same pace (Cyranoski et al., 2011; Sauermann and Roach, 2012). This trend has determined a fierce competition for job positions available after the Ph.D. (Freeman et al., 2001). A recent article in *Nature career news* surveying 317 early-career researchers seeking academic positions warned students who want to undertake an academic career that at least 15 job applications are needed to receive a single job offer (Fernandes et al., 2020; Notman and Woolston, 2020). The hyper-competition in the job market requires Ph.D. students to focus on the outcomes with high value for recruiters, who select candidates showing a solid publication profile and a rich scientific network (Alberts et al., 2014). Despite the call from the scientific community to give less weight to publication metrics in the selection decisions (Benedictus et al., 2016), the practice of publication and citation counting persists, and the norm for Ph.D. students is to publish their thesis chapters even before graduation (Black and Stephan, 2010; Brischoux and Angelier,

* Corresponding author.
E-mail address: alberto.corsini@univ-cotedazur.fr (A. Corsini).

[1] https://www.findaphd.com/advice/blog/4554/the-best-thing-about-my-phd-supervisor-students-share-their-stories.
[2] https://www.ese.wustl.edu/~nehorai/students/testimonials.html.

https://doi.org/10.1016/j.respol.2022.104561
Received 1 July 2021; Received in revised form 21 March 2022; Accepted 21 May 2022
Available online 7 June 2022

2015; Horta and Santos, 2016; Sauermann and Haeussler, 2017; van Dijk et al., 2014). In addition to the publication record, recruiters also value the candidate's scientific network (Heffernan, 2021) as a signal of the candidate's capacity to establish research collaborations in an era in which knowledge production is increasingly the result of a team effort (Katz and Martin, 1997; Wuchty et al., 2007). Therefore, students, directors of Ph.D. programs, and policymakers aiming at graduating students highly competitive on the job market urge to understand which working conditions are associated with students' high publication scores and large scientific networks. In other words, they urge replying to the question: what makes a Ph.D. student productive?

In this paper, we address this question by analyzing the role that a broad range of characteristics of the social environment to which a Ph.D. student is exposed has on the student's scientific productivity during the Ph.D. period. Doing so, we provide three contributions to the extant literature.

First, extant studies show scattered empirical evidence on how factors such as university quality, supervisor's gender, supervisor's scientific network, student's nationality, students group specialization, and funding relate with students' productivity (Baruffaldi et al., 2016; Conti et al., 2014; Gaulé and Piacentini, 2013, 2018; Horta et al., 2018; Pezzoni et al., 2016; Rossello et al., 2020; Waldinger, 2010). Our paper encompasses in a unique analysis a comprehensive set of relevant biographic and academic characteristics of the supervisor and peers, the two most important actors with whom the student establishes relationships during the training period (Carayol and Matt, 2004; Shibayama et al., 2015; Stephan et al., 2007; Stephan and Levin, 1997).

Second, although the student-supervisor relationship has already received attention (Paglis et al., 2006; Platow, 2012; Sinclair et al., 2014), some relevant supervisor characteristics have been neglected by the past empirical literature, such as supervisors' mentorship experience and fundraising ability. Moreover, the empirical literature has often overlooked the influence of peers' characteristics during the student's training period (with some notable exceptions such as Broström, 2019). Nonetheless, students spend most of their time in labs, frequently interacting with their peers, making group dynamics fundamental for the student's learning process (Shibayama and Kobayashi, 2017). Our paper contributes to advance knowledge on how neglected supervisor characteristics and peers' characteristics are associated with the student's productivity.

Third, extant studies on students' productivity rely on selected samples of students affiliated to top-tier universities (Pezzoni et al., 2016), working in specific disciplines (Delamont and Atkinson, 2001; Gaulé and Piacentini, 2018), or graduating in specific years (Broström, 2019; Shibayama and Kobayashi, 2017). Using selected samples is sometimes highly desirable allowing for solid identification strategies (Waldinger, 2010). However, it comes at the cost of limited external validity of the results of the analyses due to the necessity of drawing conclusions for specific disciplines, universities, or historical periods. Our analysis overcomes this limitation by covering all French universities in all STEM fields over a long time span, including 15 cohorts of students.

The study most similar to ours is a paper by Broström (2019). Brostöm investigates how department conditions relate to Ph.D. students' early career success. He employs data on Swedish students and finds that they perform better in the early stages of their careers when trained in small teams and supervised by a professor with a solid academic profile. A key difference from our work is that Brostöm looks at a selected sample of surveyed students who graduated from one cohort and work in group, while we use data on the entire population of one country, including 15 cohorts of graduate students. Having the whole population, our study does not suffer the representativeness concerns of survey data, and the long-time span observed allows us to control for cohort effects. Another important difference from our work is that Brostöm investigates the relationship between the Ph.D. environment and postgraduation outcomes. Former Ph.D. students might have entered very different job contexts, with some students working in highly reputed universities after graduation while others quitting academia. The postgraduation environment might drive part of the identified effects. In contrast, our work bounds Ph.D. students' productivity during their training period, associating the social environment with outcomes strictly related to the training period. Our results are, therefore, informative of the effectiveness of Ph.D. programs.

Our results show that higher student productivity is associated with having a productive female supervisor. On the contrary, having a supervisor with long mentoring experience and a supervisor in early- or late-career phases is associated with lower student productivity. The supervisor's fundraising ability at the national and European level is associated with higher visibility of the Ph.D. student's work, as shown by the citations received by the student's Ph.D. publications. However, being supervised by a researcher awarded European grants negatively relates to the student's publication quantity and network size. Looking at peers' characteristics, we find that a high number of peers is associated with a lower student's productivity. Conversely, having freshman highly-cited peers is positively associated with the student's productivity. When we break down our analysis by field, i.e., Mathematics, Engineering, Physics, and Medicine-biology-chemistry, we find heterogeneous results across fields.

2. Social environment and productivity during the training period

Understanding how the social environment characteristics are associated with the productivity of Ph.D. students is under the spotlight in the current discussion within the scientific community (Chenevix-Trench, 2006; Lempriere, 2020). This discussion has become of primary importance due to the sharp rise in the number of Ph.D. holders in the last decades that have not corresponded to an equal rise in the number of research job positions (Cyranoski et al., 2011). The mismatch between the supply and demand of Ph.D. graduates has strengthened the competition for the few available positions (Brischoux and Angelier, 2015; Freeman et al., 2001; Mangematin, 2000; van Dijk et al., 2014). When asked, more than one-third of French Ph.D. students declare nowadays to be worried about their professional future (Pommier et al., 2022). In the hyper-competitive context created, Ph.D. graduates aiming at pursuing a research career are mainly evaluated on their publication outcomes during their training period and their collaboration network (Alberts et al., 2014; Heffernan, 2021; Mangematin and Robin, 2003). Therefore, publication outcomes and collaboration networks have become fundamental assets in determining graduates' career success (Allison and Stewart, 1974; Long and McGinnis, 1985; Merton, 1968; Tenenbaum et al., 2001; Vale, 2015).

As in any other working context, students' achievements measured by publication outcomes and collaboration networks depend on the characteristics of the environment to which students are exposed. During the Ph.D. training, the relevant environment for academic training is the lab where a student works (Shibayama and Kobayashi, 2017). Within the lab, students develop social relationships with their supervisors and peers. The successful completion of the Ph.D. and the graduate program satisfaction depend on these relationships (Lovitts, 2001; Tompkins et al., 2016). Therefore, we expect supervisor and peers' characteristics to be associated with the student's productivity.

2.1. Supervisor's characteristics and student's productivity

Student's productivity largely depends on the success of the student-supervisor collaboration. As in any other collaboration relationship, collaborators' characteristics play a crucial role in determining the success of the collaboration. In the case of scientific collaborations, biographic characteristics and academic profile are essential elements to consider (Azoulay et al., 2010; Bercovitz and Feldman, 2011; Bozeman and Corley, 2004; Katz and Martin, 1997; Lee and Bozeman, 2005; Taylor and Greve, 2006). Even more so, in the student-supervisor collaboration where the supervisor's characteristics are expected to be crucial for the student's productivity due to mentorship and lab leadership role played by the supervisor (Delamont and Atkinson, 2001; Golde, 2005; Lee et al., 2007; Lempriere, 2020; Liénard et al., 2018; Ma et al., 2020; Mangematin and Robin, 2003; Pearson and Brew, 2002; Shibayama, 2019; Shibayama et al., 2015; Tenenbaum et al., 2001).

Looking at the supervisor's biographic characteristics, we consider gender, seniority, and mentorship experience. We expect female and male supervisors to have different mentorship approaches. Ethnographic studies investigating the lab routines have explored these behaviours. Surveying 185 students at the University of California, Tenenbaum et al. (2001) find that male supervisors are less likely than their female counterparts to provide psychological help to the students decreasing students' level of satisfaction with the Ph.D. training experience. However, female and male supervisors offer equal “instrumental help,” providing students the same technical knowledge needed to support their publication productivity. In another survey study in medicine, Luckhaupt et al. (2005) find that female supervisors perceive gender-related boundaries in collaborating with their students. In a field experiment involving the recruitment of students in lab manager positions, 127 professors evaluating students' resumes have been funded to favor male students (Moss-Racusin et al., 2012). Larger empirical studies have confirmed those differences. In a sample of 20,000 U.S. Ph.D. graduates in chemistry, Gaulé and Piacentini (2018) find that students pairing with a same-gender advisor are more productive than students working with an advisor of a different gender. Similar results have been found for South African Ph.D. students by Rossello et al. (2020), who show that female students working with male supervisors are less productive than male students. In the context of a leading US interdisciplinary university, Pezzoni et al. (2016) find that having a female supervisor increases Ph.D. students' productivity.

Another supervisors' biographic characteristic that is expected to affect students' productivity during the training period is the supervisors' seniority. A rational individual decreases the working time with seniority (Diamond, 1984; Levin and Stephan, 1991). In the case of scientists, we expect that they allocate their time differently across activities as seniority increases. Indeed, scientists have a high degree of autonomy in choosing the time to allocate to different activities such as fundraising, research, teaching, consulting, and administrative activities (Libaers, 2012; Sabatier et al., 2006). We expect that young supervisors aiming to boost their careers devote more time to fundraising, research, and mentoring activities. In contrast, senior supervisors are likely to dedicate more time to remunerative activities in the short term, such as consulting and administrative activities. Consequently, the less time spent in research and mentoring activities by a senior supervisor might negatively impact the support provided to her Ph.D. students, and ultimately on her students' productivity.

While seniority is expected to harm students' productivity, we expect a positive relationship between the supervisor's mentorship experience and student's productivity. Accumulated experience in supervising students develops different abilities, such as advising, tutoring, encouraging, providing a role model, and conveying to students technical and tacit knowledge (Broström, 2019; Overington, 1977). Therefore, the supervisor's mentoring skills are expected to evolve with experience and lead to better student training when the supervisor has a long mentoring history. This better training is expected to be associated with the higher productivity of the Ph.D. student during the Ph.D. period.

Looking at the supervisor's academic profile, we consider her publication record, scientific network, and fundraising abilities. Publications and citations received reflect the supervisor's academic status and scientific competencies. Ph.D. students supervised by highly productive scientists are expected to acquire practical knowledge on how to conduct successful research (Long and McGinnis, 1985; Sinclair et al., 2014). Indeed, the supervisor often becomes a model for the student who reproduces the same successful research methodologies, develops similar skills and competencies, and applies the same commitment to research enterprises (Paglis et al., 2006). Mimicking a productive supervisor's successful behaviour is expected to increase the student's productivity during the Ph.D. period.

The dimension of supervisors' network is also expected to be associated with students' productivity. For example, students supervised by scientists in contact with many co-authors are expected to be more likely to spend visiting periods in other labs acquiring new competencies, be introduced to leading scientists in the discipline, and be exposed to different research approaches (Mangematin and Robin, 2003; Stephan, 2006). These networking opportunities are expected to positively impact students' productivity (Lee and Bozeman, 2005).

Besides publication and networking influence, supervisors are also fundamental in providing resources that contribute to the students' Ph.D. program completion. Scholars have focused on assessing the role played by different types of scholarships on students' productivity (Horta et al., 2018). However, modern labs have ‘firm-like’ characteristics (Etzkowitz, 2003), making their competitiveness and survival substantially dependent on the amount of funds the professor leading the lab can raise (Stephan, 2012). Supervisors' fundraising activity is essential to support students' conference participation, visiting periods in other research institutes, and access to up-to-date lab equipment. Therefore, the supervisor's abundance of research funding is expected to be positively related with the Ph.D. student's productivity during the training period.

2.2. Peers' characteristics and student's productivity

Our study considers the student's peers as the other students exposed to the same work environment, i.e., having the same supervisor as the focal student during the same training period (Conti et al., 2014; Luckhaupt et al., 2005).

Ph.D. students, like any other worker, interact with peers during their professional activities. These interactions might affect students' productivity in several ways. First, students feel the “peer pressure” of maintaining a level of productivity similar to that of their peers striving for scientific recognition from their supervisor and the scientific community (Stephan and Levin, 1992). Moreover, the comparison with productive peers triggers psychological mechanisms of social comparison, making the focal student adopting the same productive behaviours as her colleagues (Tartari et al., 2014). Finally, students learn by observing and interacting with their peers stimulating the generation of novel research ideas (Ayoubi et al., 2017; Cornelissen et al., 2017; Delamont and Atkinson, 2001). Although peer pressure and learning from peers' mechanisms are expected to increase the student's productivity during the training period, coordination costs and competition dynamics might be detrimental to large groups' productivity (Broström, 2019).

The labour literature, both using observational and experimental data, is convergent in showing that having peer co-workers in the work environment positively affects productivity (Falk and Ichino, 2006). However, we expect the beneficial effect of having peers shrinking when the peers' number increases (Shibayama and Kobayashi, 2017). Indeed, the supervisor's time allocated to each student might reduce when the number of students increases, and the upsurge of competitive dynamics between peers might discourage students' collaboration (Conti et al., 2014).

Not only the mere presence of peers affects the focal student's productivity, but also peers' characteristics. Similar to the supervisor, we analyze peers' biographic and academic characteristics.

As biographic characteristics, we expect that both the gender and seniority of peers are associated with students' productivity. Previous studies have not reached convergent results on gender. Looking at undergraduate students, Dasgupta et al. (2015) find that group dynamics are not gender-neutral. For instance, female students' participation and self-confidence in group discussions are higher in female-majority groups. Looking at Ph.D. students, Pezzoni et al. (2016) found that, although student and supervisor's gender matters, the gender composition of the lab is not associated with the Ph.D. student's productivity. Regarding the peers' seniority, having more senior peers with greater knowledge stocks is expected to enhance knowledge transfer toward the focal student (Ayoubi et al., 2017; Delamont and Atkinson, 2001), and it might increase students' productivity. However, more senior peers might be in a phase of their Ph.D. when ideas are already settled, and interacting with other students might be less fruitful.

As peers' academic characteristics, we consider peers' publication and citation productivity. Previous literature has shown that peers' productivity positively affects individuals' productivity for low-skilled jobs such as supermarket workers and fruit-pickers (Bandiera et al., 2009; Mas and Moretti, 2009). For high skilled jobs, such as scientific research, results are not convergent. While Azoulay et al. (2010) show a decrease in the scientific productivity of team members when the team "star scientist" dies, Waldinger (2012) finds no effect of losing a brilliant peer. Although these not convergent results, in the Ph.D. students' context, we expect that highly productive peers will benefit the focal student's productivity through the three mechanisms described above: "peer pressure" adoption of productive behaviours inspired by peers through the mechanism of social comparison, and enhanced probability of acquiring knowledge from productive peers.

The mechanism of social comparison might also play a role in encouraging the expansion of the focal student's network. Although we have argued that students mainly rely on their supervisor's network to create their collaboration network, students surrounded by peers who invest energies in developing their co-authorship network during conference participation and visiting periods probably will tend to mimic the same behaviour. Therefore, we expect the student's network size to be larger when peers have a larger network.

3. The French population of STEM Ph.D. students

Our empirical setting is represented by the entire population of STEM Ph.D. students of one European country, France. The excellence of France in STEM fields is proved by the worldwide recognition gained by its scholars and top-tier research institutions. Looking at the absolute number of Nobel Prize winners, 39 French scientists obtained the highest recognition in Chemistry, Medicine, and Physics. A French elite institute, the *École Normale Supérieure in Paris*, is ranked first together with the *California Institute of Technology* by the proportion of alumni who obtained the prize. Marie Curie, the first woman who obtained a Nobel Prize and the only woman awarded twice, received her training mainly in Paris, where she established her lab. France does exceptionally well also in Mathematics, being one of the top-5 countries for the number of Fields medals.

In training scientists, France has a well-structured doctoral offer. Ph. D. scholarships are sponsored by universities, laboratories, the State, or private companies. Students are supported by scholarships that usually last three years (Pommier et al., 2022).[3] Students' hiring contracts are relatively standard, and almost all students are hired as full-time professional researchers for the entire duration of their Ph.D. (Mangematin, 2000). A centralized system standardizes doctoral program regulations, but each university has margins of flexibility in organizing courses and lab activities. Usually, programs show field heterogeneity. For instance, Ph.D. students in natural and technological sciences work full time in research labs with their colleagues, while in the other disciplines, students' work does not require a daily presence in labs. During their first year, Ph.D. students are often asked to attend core classes in theory and methodology and additional skill classes such as "writing scientific papers". In later years, a considerable amount of a student's Ph.D. time is devoted to writing the thesis, a document of about 200 pages where the student proves her research abilities. The prevalent thesis format has evolved over time, from producing a coherent monography on a specific subject to the current standard of producing a collection of three independent research articles. This change is in line with the attempt to encourage young scholars to publish their Ph.D. research work in scientific journals to facilitate their future careers. The final thesis importance is evident from the fact that French researchers often interchange the expression "being enrolled in a Ph.D. program" with "*faire une these*" (the English equivalent of "writing a thesis"). Candidates need to be paired with a thesis supervisor who accepts to guide them to access the doctoral program. The practice of writing a thesis under the guidance of a supervisor assisted by a co-supervisor is allowed.

4. Data sources

To construct our study sample, we gather data from multiple sources. The first is the French repository of *Electronic Doctoral Theses*. By special permission, we obtained access to the whole universe of STEM thesis records collected by the *Agence Bibliographique de l'Enseignement Supérieur* (ABES) that is managing the repository since 1985. For each thesis record, we have information about author, abstract, university of graduation, defense date, supervisor's name, co-supervisor's name (if any), and field of study. As fields, we distinguished theses in Mathematics, Engineering, Physics, and Medicine-biology-chemistry.[4] Unfortunately, the records do not report the student's year of entry into the Ph. D. program; thus, we approximate it assuming that each student started the program three years before her thesis defense year. According to the national statistics for STEM fields, the most frequent duration of the Ph. D. training in France is four years, three years plus the thesis defense year.[5] Hence, we set the student's *entry year* into the Ph.D. program in year *t-3*, and we define the *Ph.D. training period* as the period ranging from *t-3* to *t*, where *t* is the defense year.

Our information on the students' and supervisors' gender results from a multiple-iteration matching strategy (Gaulé and Piacentini, 2018; OECD, 2012). First, we match the students' given names with the official French gender-name dataset.[6] Then, for the non-matched names, we

[3] In 2021, 97% of French Ph.D. students in Science and Technology fields benefitted from specific funding to support their Ph.D. training (Pommier et al., 2022).

[4] We also used a fine-grained distinction of fields based either on the Scopus field classification of supervisors' publications or on a manual attribution of the theses. The results of the fine-grained regression exercises are consistent with the ones presented in the main text. Results are available upon request.

[5] The Ph.D. duration is consistent with the duration of the scholarships. We double checked this statistic by querying a subset of universities' administration.

[6] Website: https://www.data.gouv.fr/fr/datasets/liste-de-prenoms/.

repeated the matching exercise with the *U.S. Census Bureau* gender-name dataset and the WIPO gender-name dataset,[7] respectively.

We retrieve students' and supervisors' publication records from Elsevier's SCOPUS database. We match the ABES list of students with the SCOPUS authors affiliated to French institutions using students' names and surnames as key matching criteria.[8] Similarly, we match supervisors' names and surnames with the SCOPUS authors.

We gather information on funding at the national and the European level. At the national level, we use the complete list of individual grants awarded by the *Agence Nationale de la Recherche* (ANR), the French national funding agency. Outside France, we consider the funding programs at the European level. We use the list of individual grants, *Horizon 2020* (H2020) and *Framework Programmes* (FP), awarded by the European Commission and collected in the CORDIS dataset. We match supervisors with principal investigators using their names and surnames.

To reconstruct the quality of the Ph.D. students' graduation department, we rely on the *QS university ranking.*[9] The QS university ranking provides detailed information on the universities' academic reputation at the department level and allows us to flag the top departments in each field. For instance, *Université de Paris* is in the top-20% of universities in Mathematics in France, but not in Engineering. We integrate the information from the QS ranking with bibliometric information concerning the university affiliates. We construct an appropriate bibliometric dataset of the publications and publications' authors for all the French university departments. To create this dataset, we manually match the names of the French universities (and their variants) with the SCOPUS affiliations' names. As an additional proxy for the department quality, we identify the French universities that in 2011 benefitted from the *Initiative D'Excellence* (IDEX) funding provided by the French Government to a selected group of French higher education institutions. The IDEX funding program was launched in 2011 by the French Government within a national fiscal stimulus and awarded to eight universities[10] striving to become competitors of worldwide top-ranked universities.

To create our study sample, we link all the information retrieved from the data sources listed above in a unique original dataset. Doing so, we joined student, supervisor, and department information. In addition, we refined our study sample excluding students showing productivity indicators too high to be credible.[11] Overall, the excluded students represent less than 10% of our initial sample from the ABES list of student names. After this cleaning exercise, we obtained a study sample of 77,143 Ph.D. students who graduated from French universities between 2000 and 2014.

5. Econometric methodology

To estimate how the Ph.D. student's social environment characteristics relate to her productivity, we estimate the coefficients of the model presented in Eq. (1) using *Ordinary Least Squares* (OLS). As represented by subscript *i*, the analysis is at the student level.

$$
\begin{aligned}
Student'sproductivity_i = {} & \beta_0 + (Supervisor'scharacteristics_i)'\beta_1 \\
& + (Peers'characteristics_i)'\beta_2 + (Controls_i)'\beta_3 + \varepsilon_i
\end{aligned}
\tag{1}
$$

The left-hand side variable ***Student's productivity*** in Eq. (1) takes, in turn, the value of the student's publication quantity, quality, and the size of the scientific network. We measure the publication quantity by counting the number of peer-reviewed papers published by the student (*Publications*) and the publication quality by counting the number of yearly citations received on average by the student's papers (*Average citations*). We proxy the student's research network size as the number of the student's distinct co-authors (*Co-authors*). The three productivity variables are calculated during the Ph.D. training period, i.e., from *t-3* to *t*, with the addition of one year after the thesis defense to account for possible time lags in the publication process (Powell, 2016). In other words, we calculate the productivity outcomes in the period ranging between *t-3* and $t + 1$, where *t* is the thesis defense year.

The vectors *Supervisor's characteristics* and *Peers' characteristics* define the Ph.D. student's social environment. *Controls* is a vector including the student's characteristics and the characteristics of the department where the student is enrolled. Finally, ε is the idiosyncratic error term. Our interest is to estimate the vectors of coefficients β_1 and β_2 that relate supervisor and peers' characteristics with the student's productivity.

A concern in estimating these coefficients regards a potential endogeneity issue. Although we include in our regression a large set of time-variant and time-invariant characteristics identified by the previous literature as factors affecting the student's productivity, the lack of proxies for the student's intrinsic ability might bias our estimates. Indeed, an omitted variable problem might arise if the unobserved ability correlates with explained and explanatory variables. For instance, students with higher research ability might be at the same time more productive and more likely to be supervised by scientists with better academic credentials. However, previous studies have shown that this endogeneity problem is mitigated by the supervisor's difficulty in assessing the student's research ability when the student is at the beginning of her academic career (Mangematin, 2000). The asymmetry of information during the student's selection process makes it unlikely to observe a correlation between students' intrinsic ability and supervisors' quality. Belavy et al. (2020) show in an empirical study on 324 Ph.D. students that variables usually used as proxies for the students' ability, such as previous academic achievements and training grades, are uncorrelated with the students' Ph.D. productivity. Along the same line, anecdotal evidence shows that standardized tests often considered for Ph.D. enrollment, e.g., GRE scores in the U.S., do not fully reflect the student's future academic ability (Aristizábal, 2021). Although previous literature excludes a strong correlation between the student's academic ability and the supervisor's quality, in Appendix E, we implement a robustness check to respond to the potential endogeneity concern. Specifically, we replicate the estimations of Eq. (1) adding a proxy that controls for the ability of the student during her high school period. We flag students with exceptional ability by calculating a dummy variable equal to one if the student has participated in a selective contest during high school (Agarwal and Gaule, 2020). We consider three well-known contests: the *International Mathematical Olympiad* (IMO), *Les Olympiades Nationales de Mathématiques* (the national French Mathematical Olympiad), and *le Kangourou des mathématiques* (a French national mathematical contest). We find that including a proxy for the student's ability does not affect the estimated coefficients of the variables in the *Supervisor's characteristics* and *Peers' characteristics* vectors, showing that our results are unlikely to be affected by an endogeneity problem.

5.1. Supervisor's characteristics

We consider the supervisor's biographic and academic characteristics. As for the biographic characteristics, we include a dummy variable

[7] Website: https://www.wipo.int/publications/en/details.jsp?id=4125.

[8] We dropped from the initial list of students provided by ABES students with homonymous names. Having two or more students with the same full name in our original list of Ph.D. thesis authors would make it difficult to disentangle their identity and correctly assign bibliometric information. Therefore, we decided to drop the homonyms from our original list of Ph.D. thesis authors.

[9] Website: https://www.topuniversities.com.

[10] The 8 awarded universities are: Université d'Aix-Marseille, Université de Bordeaux, Université Paris Saclay, PSL Paris Sciences et Lettres, Sorbonne Université, Sorbonne-Paris-Cité, Université de Strasbourg, Université de Toulouse.

[11] We excluded students with more than 20 publications, more than 100 citations received per paper, and more than 200 co-authors during the Ph.D. period. We excluded also students for which their supervisors reported more than 100 publications and more than 500 co-authors during the five years preceding the student enrollment.

Female supervisor that equals one if the supervisor is a female scientist, zero otherwise. Expecting that the attention dedicated to a Ph.D. student varies along the supervisor's career, we calculate the *Supervisor's seniority* measured as the years elapsed between the supervisor's first publication and the student's entry year into the Ph.D. program. To capture possible nonlinear effects of seniority, we include a squared term of the variable *Supervisor's seniority*. Also, the mentorship experience of the supervisor might affect the productivity of her Ph.D. students. Therefore, we calculate the variable *Mentorship experience* as the cumulated number of students mentored by the supervisor who have successfully defended their thesis until the focal student's entry year into the Ph.D.[12]

Concerning the supervisor's academic characteristics, we calculate two variables proxying the supervisor's publication quantity and quality in the five years preceding the entry of her student into the Ph.D. program, i.e., from *t-8* to *t-4*, where *t* is the student's defense year. We decided to measure the supervisor's publication quantity and quality during the five years preceding the student enrollment (and not during the student training period) because it is a common practice that the student and her supervisor co-sign articles during the student's training period. In the case of co-signed articles, it is impossible to disentangle supervisor and student's productivity. We define the variable *Supervisor's publications* as the number of supervisor's publications in peer-reviewed journals over the five years preceding the student's entry into the Ph.D. program. Then, for the same period, we calculate the average number of yearly citations received by the supervisor's articles (*Average citations*). To proxy for the supervisor's scientific network size, we reconstruct her co-authorship network. We define the variable *Supervisor's co-authors* as the number of distinct co-authors that the supervisor has in the five years preceding the student's entry into the Ph.D. program. Finally, to proxy for the supervisor fundraising ability, we calculate a dummy variable *ANR grant* that equals one if the supervisor is the principal investigator of an ANR grant in at least one year of the student's training period. Similarly, we define a dummy variable *EU grant* that equals one if the supervisor is the principal investigator of at least one EU grant during the student's training period.

5.2. Peers' characteristics

Ph.D. students might spend their Ph.D. training period alone if their Ph.D. period does not overlap with the Ph.D. period of other students. In the opposite case, they might share the Ph.D. experience with other peer students. To distinguish these two cases, we calculate the dummy variable *With peers* that takes value one if the focal student spends at least one year of her training period with at least another student having the same supervisor, zero otherwise. Then, we calculate the variable *N. peers* as the average yearly number of students with whom the focal student shares the training experience. Students start their Ph.D. training in different moments, and cohorts of students can overlap only partially. To calculate the variable *N. peers*, we first calculate the yearly number of peers in each of the four years of the focal student's training period; then, we average the four values. For instance, if the focal student spends the first three years alone and her supervisor recruits another student during her last Ph.D. year, the variable *N. peers* equals 0.25 ($0.25 = (0 + 0 + 0 + 1) / 4$).

To characterize the relationships between the student and her peers, we calculate variables proxying for the peers' biographic and academic characteristics. Concerning the biographic characteristics, we calculate the dummy variable *At least one female peer* that equals one if at least one peer during the focal student's training period is a female student, zero otherwise. We also calculate the peers' average seniority as the average number of years spent by the peers in their Ph.D. program (*Average peers' seniority*). Also, in this case, peers might have training periods that only partially overlap with that of the focal student. Thus, as the first step of the peers' seniority variable construction, we calculate the average peer seniority in each year of the 4-years of the focal student's training period. If the focal student has no peers in one year, we assign the value zero to the average yearly seniority. Then, we obtain the *Average peers' seniority* variable averaging the four yearly values. For instance, if the focal student has only one peer during the first two years of her training period, it means that the peer defended her thesis during the focal student's second year of Ph.D. Thus, we consider the peer's seniority values for the first two years of the focal student's training equal to 3 and 4. The variable *Average peers' seniority* equals 1.75 ($1.75 = (3 + 4 + 0 + 0) / 4$) for the focal student.

Concerning the academic characteristics, we calculate the peers' number of publications per year (*Peers' publications*). This variable is calculated following a two-step procedure. In the first step, we count the number of articles published by the peers in each of the four years of the focal student's training period. In case the focal student has no peers in one year, we assign the value zero to the yearly number of articles published. Then, we obtain the *Peers' publications* by averaging the four values. For instance, if the focal student has two peers who publish one article each[13] during the first year of her training period, the value of *Peers' publications* equals 0.5 ($0.5 = (2 + 0 + 0 + 0) / 4$). Applying the same two-step procedure as for the *Peers' publications*, we calculate the variable *Peers' average citations* proxying for the quality of peers' work and the variable *Peers' co-authors* proxying for the peers' network size.

5.3. Other controls

To mitigate a potential bias of our estimated coefficients, we control for the department and student's characteristics. We define a department as the pair university-field. For instance, *Université de Paris* counts four departments: *Université de Paris*-Mathematics, *Université de Paris*-Engineering, *Université de Paris*-Physics, and *Université de Paris*-Medicine-biology-chemistry.

To control for department quality, we retrieve the university reputation ranking from the QS World University ranking.[14] We create a dummy *French Top-20* that equals one if the department is among the 20% of departments with the highest academic reputation in a specific field in France. As an additional proxy for the department quality, we calculate the average citation-weighted publication productivity per department affiliate (*Citation-weighted publications per affiliate*). To calculate this latter variable, we consider the department affiliates' average productivity during the five years preceding the student's entry into the Ph.D. program. Specifically, we identify the department affiliates' publications during the five years preceding the student enrollment; then, we weigh each publication by the citations received each year. Finally, we calculate the average number of affiliates' citation-weighted publications for each department. We also calculate the variable *IDEX* as a third control for the department quality. This variable is a dummy that equals one after 2011 if the student's department was selected to be awarded the IDEX national investment program funding.

To control the department size, we calculate the variable *Department size* counting the number of scientists affiliated to the department for at least one year during the five years preceding the student's entry into the Ph.D. program. We rescale the number of affiliates dividing by 100, meaning that each unit increase of the variable *Department size*

[12] We retrieve data on supervisors' mentoring career starting from 1980.

[13] In case of joint publications between two or more peers of the same focal Ph.D. student, we count the publication once.

[14] https://www.topuniversities.com/university-rankings. We gather the ranking information in 2020, however university ranking has minor variation over the years when considering top-universities. The advantage of using the QS World University ranking is the availability of a ranking that is detailed by subject area.

corresponds to 100 additional department affiliates.[15]

Along with the department size, the size of the Ph.D. program might also play a role. Larger Ph.D. programs might be better organized and provide students with a better and more productive training experience. We calculate the number of Ph.D. students enrolled in the same focal student's Ph.D. program for each of the four years of her training period. Then, we calculate the variable *N. of Ph.D. students in the program* averaging the four yearly values.

Finally, we control for the characteristics of the Ph.D. student. Specifically, we control for the gender of the student with a dummy variable *Female student* that equals one for female students, zero otherwise.[16] We consider the student's possibility of having a thesis co-supervisor defining the dummy *Co-supervision* that takes value one in the presence of a co-supervisor, zero otherwise. We also add four dummy variables, *Mathematics*, *Engineering*, *Physics*, and *Medicine-biology-chemistry* controlling for the heterogeneity across the thesis research fields. Finally, we add a set of dummy variables for the students' *Entry year* into the doctoral program to account for the Ph.D. cohort effect.

5.4. Descriptive statistics

Table 1 lists all the variables included in our analysis with a short description. Table 2 reports the descriptive statistics for the variables calculated on our sample of 77,143 Ph.D. students. When classified by field, 15% of the students are in Mathematics, 18% in Physics, 21% in Engineering, 45% in Medicine, Biology, and Chemistry. Students publish on average 2.37 peer-reviewed articles during their training period. 68% of students publish at least one article during the Ph.D. period. The average students' collaboration network includes 8.93 distinct co-authors.

The average supervisor has a stock of 13.59 peer-reviewed articles and a seniority of 11.49 years of career when her student enrolls in the Ph.D. program. At the time of the student enrollment, the average supervisor counts 3.08 successfully supervised Ph.D. students over her career. For the gender composition, 39% of Ph.D. students are women, while this percentage reduces to 21% when looking at the supervisors. Only 6% of the students have a supervisor who is the principal investigator of an ANR national grant during the Ph.D. training period, and only 2% of the students have a supervisor who is the principal investigator of a EU grant.

Looking at the focal Ph.D. student's peers, 80% of the students have at least one peer during the training period, and, on average, they are in contact with 1.76 peers per year. During the training period, the focal student's peers publish on average 0.81 papers per year.

Table A1, in Appendix A, reports the variable correlation matrix.

6. Results

Table 3 reports the OLS estimates of the model described in Eq. (1).

Looking at the impact of the biographic characteristics of the supervisor on the student's productivity, we find that having a *Female supervisor* is not associated with the number of papers published by the student. On the contrary, having a female supervisor is associated with a higher number of citations (+0.074 yearly citations per paper) and a larger collaboration network (+0.31 co-authors). These two variations are statistically significant and have economic relevance, corresponding

[15] In an alternative model specification, we include department fixed effects. Our results are unchanged and available upon request.

[16] We do not have information about the age of the Ph.D. students, however in France students tend to enroll in the Ph.D. program soon after their master studies, thus we do not expect much age heterogeneity among students.

Table 1
List of variables used in the analysis.

	Variable description
Dependent variables	
Student's productivity	
Publications	Ph.D. student's number of papers published between t-3 and t + 1[a]
Average citations	Average yearly citations received by the student's papers published between t-3 and t + 1
Co-authors	Number of distinct co-authors of the student between t-3 and t + 1
Independent variables	
Supervisor characteristics	
Female supervisor	Dummy variable that equals one if the supervisor is a female scientist
Supervisor's seniority	Number of years elapsed from the first supervisor's publication to t-3
Mentorship experience	Cumulated number of Ph.D. students successfully supervised until t-3
Supervisor's publications	Supervisor's number of papers published between t-8 and t-4
Supervisor's average citations	Average yearly citations received by the supervisor's articles published between t-8 and t-4
Supervisor's co-authors	Supervisor's number of distinct co-authors between t-8 and t-4
ANR grant	Dummy variable that equals one if the supervisor is the principal investigator of an ANR grant between t-3 and t
EU grant	Dummy variable that equals one if the supervisor is the principal investigator of a EU grant between t-3 and t
Peer characteristics	
With peers	Dummy variable that equals one if the student has at least one peer between t-3 and t
N. peers	Average number of the student's peers per year between t-3 and t
At least one female peer	Dummy variable that equals one if at least one student's peer is a female student between t-3 and t
Average peers' seniority	Average yearly seniority in the Ph.D. program of the student's peers
Peers' publications	Average number of peers' publications per year between t-3 and t
Peers' average citations	Average yearly citations received by the peers' articles between t-3 and t
Peers' co-authors	Peers' average number of distinct co-authors per year between t-3 and t
Other controls	
French Top-20	Dummy variable that equals one if the student's department is among the 20% departments with the highest academic reputation score in France according to the QS ranking
Citation-weighted publications per affiliate	Average department affiliate's citation-weighted publication productivity between t-8 and t-4
IDEX	Dummy variable that equals one if t is greater or equal to 2011 and the student is enrolled in a university awarded IDEX funding
Department size [100 affiliates]	Total number of scientists affiliated to the student's department between t-8 and t-4
N. of Ph.D. students in the program	Average number of Ph.D. students per year enrolled in the focal student's Ph.D. program between t-3 and t
Female student	Dummy variable that equals one if the Ph.D. student is female
Co-supervision	Dummy variable that equals one in the presence of a co-supervisor
Mathematics	Dummy variable that equals one if the Ph.D. dissertation is in Mathematics
Engineering	Dummy variable that equals one if the Ph.D. dissertation is in Engineering
Physics	Dummy variable that equals one if the Ph.D. dissertation is in Physics
Medicine-biology-chemistry	Dummy variable that equals one if the Ph.D. dissertation is in Medicine, Biology, or Chemistry
Entry year	The student's entry year into the Ph.D. program, i.e., t-3

[a] t is the Ph.D. thesis defense year; t-3 is the entry year of the student into the Ph.D. program; the four years ranging from t-3 to t define the Ph.D. training

period; the five years ranging from t-8 to t-4 are the years preceding the student's entry into the Ph.D. program.

Table 2
Descriptive statistics for our sample of 77,143 Ph.D. students.

77,143 Ph.D. students	Mean	SD	Min	Max
Dependent variables				
Ph.D. student				
Publications	2.37	2.99	0.00	20.00
Average citations	2.11	3.51	0.00	98.14
Co-authors	8.93	15.37	0.00	200.00
Independent variables				
Supervisor characteristics				
Female supervisor	0.21	0.41	0.00	1.00
Supervisor's seniority	11.49	5.24	0.00	21.00
Mentorship experience	3.08	6.22	0.00	184.00
Supervisor's publications	13.59	14.31	0.00	100.00
Supervisor's average citations	2.36	3.03	0.00	127.87
Supervisor's co-authors	37.28	50.82	0.00	499.00
ANR grant	0.06	0.25	0.00	1.00
EU grant	0.02	0.16	0.00	1.00
Peer characteristics				
With peers	0.80	0.40	0.00	1.00
N. peers	1.76	2.14	0.00	30.00[a]
At least one female peer	0.52	0.50	0.00	1.00
Average peers' seniority	1.61	1.04	0.00	3.56
Peers' publications	0.81	1.76	0.00	41.00
Peers' average citations	2.71	8.11	0.00	353.15
Peers' co-authors	4.21	10.28	0.00	190.75
Other controls				
French Top-20	0.39	0.49	0.00	1.00
Citation-weighted publications per affiliate	7.37	4.43	0.38	35.05
IDEX	0.18	0.38	0.00	1.00
Department size [100 affiliates]	29.25	30.28	0.04	114.46
N. of Ph.D. students in the program	1042.07	800.94	1.00	2973.00
Female student	0.39	0.49	0.00	1.00
Co-supervision	0.31	0.46	0.00	1.00
Mathematics	0.15	0.36	0.00	1.00
Engineering	0.21	0.41	0.00	1.00
Physics	0.18	0.39	0.00	1.00
Medicine-biology-chemistry	0.45	0.50	0.00	1.00
Entry year	2005.12	4.20	1997.00	2011.00

[a] Although the maximum number of peers might look high, we checked the case of the student with 30 peers during the training period. The student was supervised by a researcher in Physics, having yearly 30(+1) Ph.D. students during the focal student's training period.

to the 3.5%[17] of the sample average student's citations and 3.5% of the sample average student's co-authors. Regarding the *Supervisor's seniority*, we find an inverted U-shape relationship between the supervisor's seniority and the three student's outcomes considered. The maximum impact of seniority on the student's publication productivity, citations, and network size is for a mid-career supervisor, i.e., when the supervisor has 9.74,[18] 3.70, and 8.21 years of seniority, respectively.

We find that the supervisor's *Mentorship experience* is negatively associated with the student's productivity: a student mentored by an experienced supervisor shows fewer papers published, citations received, and a smaller collaboration network. Increasing by one standard deviation, the *Mentorship experience* is associated with 0.11 fewer papers,[19] 0.045 fewer citations, and 0.23 fewer co-authors. To further

[17] This percentage is calculated dividing the variation of the student's *Average citations* associated to having a *Female supervisor* by the average value of *Average citations* in the sample, reported in Table 2 (2.11).

[18] The seniority corresponding to the maximum marginal effect on publication productivity is calculated using the coefficients estimated in Column 1 of Table 3, and applying the following calculation −0.037/(2*-0.0019).

[19] The value −0.11 is obtained by multiplying the coefficient of *Mentorship experience* estimated in Table 3, Column 1, and the standard deviation of *Mentorship experience* reported in Table 2 (−0.11 = −0.018*6.22).

Table 3
Regression results. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor characteristics			
Female supervisor	−0.0051	0.074**	0.31**
	(0.025)	(0.030)	(0.13)
Supervisor's seniority	0.037***	0.0071	0.11***
	(0.0072)	(0.0085)	(0.036)
Supervisor's seniority^2	−0.0019***	−0.00096**	−0.0067***
	(0.00034)	(0.00040)	(0.0017)
Mentorship experience	−0.018***	−0.0072***	−0.037***
	(0.0019)	(0.0023)	(0.0097)
Supervisor's publications	0.027***	0.0070***	−0.10***
	(0.0012)	(0.0015)	(0.0062)
Supervisor's average citations	0.031***	0.20***	0.21***
	(0.0036)	(0.0043)	(0.018)
Supervisor's co-authors	0.0028***	0.0014***	0.091***
	(0.00034)	(0.00040)	(0.0017)
ANR grant	0.0048	0.54***	0.22
	(0.043)	(0.050)	(0.21)
EU grant	−0.19***	0.33***	−1.28***
	(0.065)	(0.077)	(0.33)
Peer characteristics			
With peers	0.13***	0.24***	0.25
	(0.041)	(0.048)	(0.21)
N. peers	−0.12***	−0.042***	−0.39***
	(0.0071)	(0.0083)	(0.036)
At least one female peer	−0.028	0.073**	0.21*
	(0.025)	(0.030)	(0.13)
Average peers' seniority	−0.14***	−0.13***	−0.63***
	(0.017)	(0.020)	(0.086)
Peers' publications	0.13***	−0.15***	−0.64***
	(0.014)	(0.016)	(0.070)
Peers' average citations	0.0065***	0.056***	0.049***
	(0.0020)	(0.0024)	(0.010)
Peers' co-authors	0.0029	0.0017	0.21***
	(0.0023)	(0.0027)	(0.011)
Other controls			
French Top-20	−0.0082	0.068**	−0.36***
	(0.023)	(0.028)	(0.12)
Citation-weighted publications per affiliate	0.012**	0.026***	0.14***
	(0.0057)	(0.0067)	(0.029)
IDEX	−0.056	0.031	−0.032
	(0.036)	(0.042)	(0.18)
Department size [100 affiliates]	0.00081	0.0014**	0.013***
	(0.00057)	(0.00067)	(0.0029)
N. of Ph.D. students in the program	0.000092***	0.00023***	0.00038***
	(0.000016)	(0.000019)	(0.000079)
Female student	−0.64***	−0.19***	−1.84***
	(0.021)	(0.025)	(0.11)
Co-supervision	−0.066***	−0.042	−0.66***
	(0.023)	(0.027)	(0.12)
Engineering	0.18***	0.40***	0.99***
	(0.035)	(0.041)	(0.18)
Physics	0.77***	0.57***	2.46***
	(0.055)	(0.065)	(0.28)
Medicine-biology-chemistry	1.54***	1.39***	6.45***
	(0.043)	(0.050)	(0.22)
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	1.23***	0.23***	3.83***
	(0.074)	(0.087)	(0.37)
Observations	77,143	77,143	77,143
R-squared	0.140	0.128	0.174

Note: Significance levels at ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are reported in parentheses. In an additional exercise, we calculate the p-values applying a multiple-inference adjustment to correct possible erroneous inferences due to the high number of hypotheses tested. Specifically, we calculate the p-values applying the Romano-Wolf multiple hypothesis correction (Romano and Wolf, 2005a,b; Romano and Wolf, 2016). The statistical significance of the coefficients remains consistent with the main results, with only two notable exceptions. Specifically, the coefficient of the variable *At least on female peer* loses its statistical significance (at standard significance levels) in the regression explaining the number of student's *Co-authors* and the coefficient of the variable *Department size* [*1000 affiliates*] loses its statistical significance in the regression explaining the *Average citations* received by the student's work. The exercise is available upon request.

investigate this result, in Appendix F, we search for non-linear relationships between *Mentorship experience* and student's productivity. Specifically, we calculate a set of dummy variables identifying different levels of experience. Fig. 1 reports the graphical representation of the marginal effects of these dummy variables as estimated in Table F1. Consistently with the results in Table 3, Fig. 1 shows that the higher the supervisor's experience, the lower the student's productivity. Focusing on the 11.7% of students supervised by researchers with a high *Mentorship experience*, i.e., researchers who supervised more than seven students before the current one, those students show 0.55 fewer publications (23.2% of the sample average), 0.25 fewer citations per paper (11.8% of the sample average), and 1.44 fewer co-authors (16.1% of the sample average) than the students supervised by mentors with no experience. This result contrasts with our expectation that being mentored by an experienced supervisor is positively associated with student's productivity. We interpret our finding as the supervisors' tendency to be more supportive to students when they are at their first experience as thesis directors.[20]

Looking at the supervisor's academic characteristics, supervisor's productivity measured by *Supervisor's publications*, *average citations*, and *co-authors*, is associated with higher student's productivity. Specifically, increasing the supervisor's publication by one standard deviation is associated with 0.39[21] additional student publications (16.3% of the sample average[22]) and 0.10 additional citations (4.75% of the sample average). Similar to *Supervisor's publications*, both the *Supervisor's average citations* and *co-authors* are associated with positive outcomes for the student along all the three dimensions considered. Increasing by one standard deviation the *Supervisor's average citations* is associated with 0.09 additional articles (3.96% of the sample average), 0.61 additional citations (28.72% of the sample average), and 0.64 additional co-authors (7.13% of the sample average). Increasing by one standard deviation the *Supervisor's co-authors* is associated with 0.14 additional articles (6.00% of the sample average), 0.07 additional citations (3.37% of the sample average), and 4.62 additional co-authors (51.79% of the sample average). The only exception to all these positive correlations is the relationship between the supervisor's number of publications and the student's network size: increasing the supervisor's publication by one standard deviation is associated with 1.43 fewer co-authors (16.02% of the sample average). This negative association might be explained by the fact that when students work with highly productive supervisors, they have fewer incentives to enlarge their network outside the lab. Despite this latter negative association, our results show a positive relationship between the supervisor's academic characteristics and the productivity of the Ph.D. student.

Considering the supervisor's fundraising ability, when the supervisor is the principal investigator of a French ANR grant, the student's work receives 0.54 additional yearly citations per paper, which corresponds to 25.59% of the students' citation average in our study sample. Similarly, having a supervisor awarded a European grant is associated with an increase of 0.33 citations received by the student's work (15.64% of the citation average). In contrast, having a supervisor awarded a European grant is associated with 0.19 fewer publications (8.02% of the publication average) and 1.28 fewer co-authors (14.33% of the co-author average). These negative correlations might be explained by the additional time spent by the supervisor managing the EU grant. Indeed, EU grants are large international projects funded by the European Commission, and supervisors need to invest a relevant amount of time in managing them. This time is probably subtracted from mentoring students. Although we observe some differences between ANR national grants and European grants, our results converge in showing that the availability of supervisor's funds is positively associated with the quality of the student's productivity.

Looking at the peers' effect, we find a positive association between the dummy variable *With peers* and the Ph.D. student's productivity. However, this variable has to be always interpreted jointly with the variable *N. peers*, since when the dummy variable *With peers* equals one, the variable *N. peers* takes positive values. For instance, we find that the overall effect of having one peer in every year of the Ph.D. period is associated with 0.20 (=0.24–0.042*1) additional citations (9.4% of the sample average), and we do not observe any statistical significance[23] of having one peer for the publication quantity and co-authorship network size. Although having one peer is associated with benefits to citations, further increasing the number of peers is associated with a decrease in all dimensions of the student's productivity, namely 0.12 fewer publications, 0.042 fewer citations, and 0.39 fewer co-authors for each additional peer.[24] These three values correspond to 5.06% of the publication average, 2.00% of the citation average, and 4.37% of the co-author average in the study sample. This empirical evidence shows that the larger the number of peers, the lower the student's productivity. Therefore, sharing the training experience with large groups of peers penalizes students' productivity, showing that the quality of the mentoring activity declines if the supervisor has many students. This decline might be related to the lack of time devoted by the supervisor to each student. Moreover, this result is particularly relevant because it suggests an optimal number of peers associated with the student's productivity. In Table F2, Appendix F, we dig into these findings to identify possible non-linear relationships between the variable *N. peers* and the student's productivity. Specifically, we calculate six dummy variables, one for each unit increase in the value of the variable *N. peers*. The alternative model specification reported in Table F2 confirms the main results reported in Table 3: having up to one peer in each year of the Ph.D. period is associated with a higher number of citations received by the doctoral student's work. On the contrary, an increase in the number of peers is associated with a decrease in all three student's productivity outcomes. Fig. 2 shows the marginal effects associated with an increasing number of peers in the student's environment on student's productivity outcomes.

[20] Interestingly, supervisor seniority is weakly correlated with the mentorship experience. This shows that, in our sample, we might observe supervisors in the early stages of their careers who accumulated a considerable mentorship experience and, vice versa, senior supervisors with no Ph.D. students. Moreover, in additional empirical analyses, we investigate the publication productivity distribution and the presence of CNRS affiliated researchers among supervisors with no previous supervision experience. We find that the publication productivity distribution for supervisors with no previous mentorship experience largely overlaps the productivity distribution of researchers with experience, meaning that there are high-quality researchers with notable publication records also among the supervisors without supervision experience. Similarly, we find that the proportion of researchers affiliated to CNRS is similar for supervisors with no supervision experience and supervisors with experience.

[21] This value is obtained by multiplying the standard deviation of the variable *Supervisor's publications* 14.31 (Table 2) by the coefficient 0.027 of *Supervisor's publications* in Table 3, Column 1.

[22] This percentage is calculated dividing the variation of the student's *Publications* associated to one standard deviation increase of *Supervisor's publications* by the sample average value of *Publications* reported in Table 2 (2.37).

[23] To test for the statistical significance of the linear combination of the coefficients of the variables *With peers* and *N. peers*, we conducted an F-test on the null hypothesis that $\beta_{With\ peers} + \beta_{N.\ of\ peers} * 1 = 0$.

[24] As a further robustness check, we run a regression selecting the subsample of 61,696 students with at least one peer. Results are consistent with those reported in Table 3.

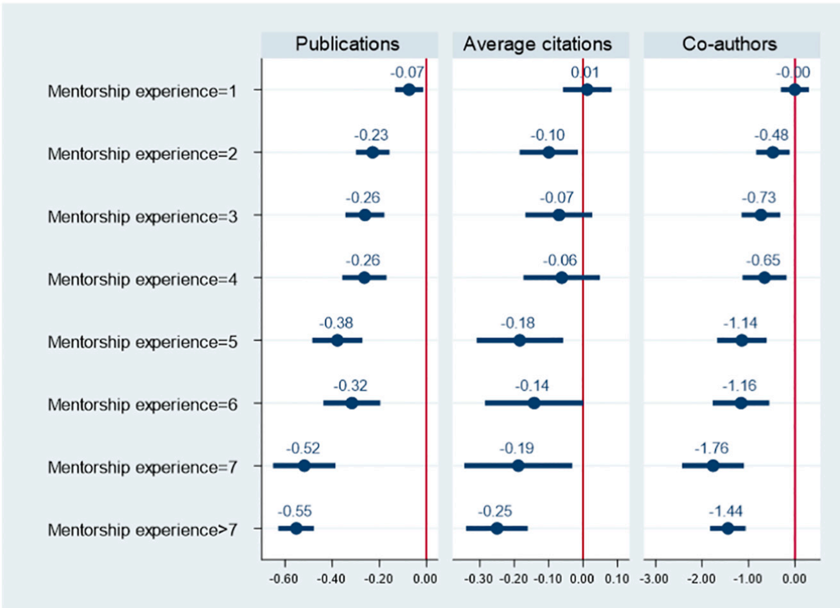

Fig. 1. *Mentorship experience* marginal effects on student's productivity outcomes.
Note: The figure reports the marginal effects estimated for the set of 8 dummy variables calculated in Appendix F and used in the regression exercises reported in Table F1. The variable *Mentorship experience* = *1*, takes value one if the supervisor has mentored only one Ph.D. student who graduated before the focal student enrollment. The variable equals one for 15.1% of the supervisors. Similarly, we calculate *Mentorship experience* = *2* (9.7%), *Mentorship experience* = *3* (7.0%), *Mentorship experience* = *4* (5.1%), *Mentorship experience* = *5* (3.9%), *Mentorship experience* = *6* (2.9%), *Mentorship experience* = *7* (2.4%), and *Mentorship experience* > *7* (11.7%). The reference case, represented by the vertical line centered in zero, is when the supervisor has *No mentorship experience* (42.1%). Bars represent 95% confidence intervals.

Conditional on having at least one peer, peers' biographic characteristics matter. Having *At least one female peer* student during the Ph.D. period is positively associated with both the focal Ph.D. student's citations received and network size, but not with the number of publications. The increase in the student's citations and co-authors equals 0.073 citations (3.46% of the sample average) and 0.21 co-authors (2.35% of the sample average). Increasing the variable *Average peers' seniority* by one standard deviation is associated with a lower focal Ph.D. student's productivity along all the dimensions considered, namely −0.15 publications (6.14% of the sample average), −0.14 yearly citations (6.41% of the sample average), and −0.66 co-authors (7.34% of the sample average). These results lead us to conclude that peers' gender positively correlates with the student's productivity, while peers' seniority negatively correlates with the student's productivity.

Regarding the peers' academic characteristics, an increase in the number of *Peers' publications* by one standard deviation is associated with fewer citations and fewer co-authors: −0.26 citations (12.51% of the sample average) and −1.13 co-authors (12.61% of the sample average). On the contrary, an increase in *Peers' publications* is associated with 0.23 additional articles published by the focal student (9.65% of the sample average). An increase of one standard deviation of the *Peers' average citations* is associated with an overall productivity boost for the focal student: +0.05 publications (2.22% of the sample average), +0.45 citations (21.52% of the sample average), and +0.40 co-authors (4.45% of the sample average). The increase of *Peers' co-authors* by one standard deviation benefits only the focal student's network size being associated with 2.16 additional co-authors (24.17% of the co-author sample average). In the light of these results, we conclude that peers' academic characteristics show mixed effects on the focal student's productivity. We can interpret these results on peers' productivity in the light of the "peer pressure" mechanism leading the student to maintain a productivity level similar to her peers. Specifically, when peers increase their publication quantity, the focal student feels the pressure to increase her outcomes in terms of quantity at the disadvantage of quality and collaboration aspects, consistently with the coefficients of the variable *Peers' publications* in the regression exercises. Differently, competition on quality between peers increases all the dimensions of scientific productivity considered, consistently with the coefficients of the variable *Peers' average citations* in the regression exercises.

For the controls, the quality of the department as measured by the variable *Citation-weighted publications per affiliate* is positively associated with all the students' productivity outcomes. On the contrary, when we measure department quality according to the variable *French Top-20*, we find that being affiliated to a top-20 reputed department positively relates to the student's citations while negatively relates to her network size. Finally, *French Top-20* is not significantly related to the number of articles published by the student. Doing a Ph.D. in a university benefitting from an *IDEX* award does not significantly correlate with the student's productivity outcomes.

The size of the department and the size of the Ph.D. student program do matter. The department size positively relates to the student's yearly citations and co-authors. Larger departments are more likely to generate internal collaborations between affiliates or attract a greater number of external collaborators. Similarly, an increase in the size of the Ph.D. program (*N. of Ph.D. students in the program*) is positively associated with all the Ph.D. student's productivity dimensions. Larger Ph.D. programs might be better structured and organized, benefitting students' productivity.

Considering the Ph.D. student characteristics, we find a significant gender gap between female and male students. Female students are less productive than their male counterparts across all the three outcomes investigated (−0.64 publications, −0.19 yearly citations, and −1.84 co-authors).[25] Moreover, the presence of a co-supervisor is associated with a decrease of the student's productivity.

Looking at the set of dummy variables identifying the fields of study, we observe productivity heterogeneity across fields. This latter result is expected since different fields are characterized by heterogeneous norms, rules, and working conditions affecting students' productivity. Following the idea that field heterogeneity matters, Section 6.1 explores the possibility of field-specific effects of our regressors by estimating the

[25] We have estimated an econometric model where we interacted the student gender with the supervisor gender. We found non-significant effects of the interaction terms. We do not report interactions in our main model specification.

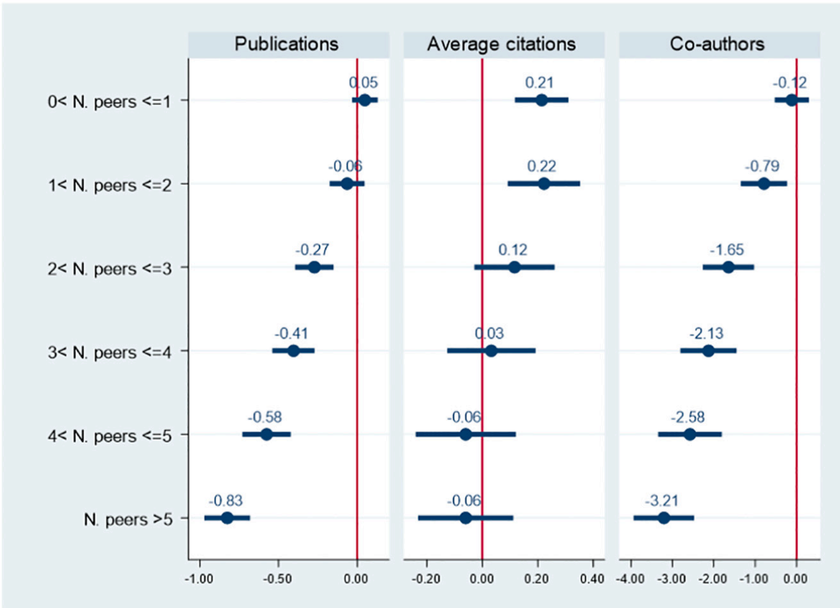

Fig. 2. *N. peers* marginal effects on student's productivity outcomes.
Note: The figure reports the marginal effects estimated for the 6 dummy variables calculated in Appendix F and used in the regression exercises reported in Table F2. The variable $0 < N.\ peers \leq 1$, takes value one if the student has between 0 (excluded) and 1 (included) peers per year during the Ph.D. period. The variable equals one for 29.9% of the students. Similarly, we calculate $1 < N.\ peers \leq 2$ (20.7% of the students), $2 < N.\ peers \leq 3$ (12.2%), $3 < N.\ peers \leq 4$ (6.9%), $4 < N.\ peers \leq 5$ (3.8%), and $N.\ peers > 5$ (6.4%). The reference case, represented by the vertical line centered in zero, is when the focal student has *No peers* (20.0%). Bars represent 95% confidence intervals.

Table 4
Ph.D. students' productivity by field.

Dependent variables	Engineering	Mathematics	Medicine-biology-chemistry	Physics
Publications	1.41	1.12	3.22	2.41
Average citations	1.27	0.88	2.96	1.97
Co-authors	4.00	2.59	13.39	8.79
Observations	16,519	11,450	35,038	14,136

coefficients of Eq. (1) for students in Mathematics, Engineering, Physics, and Medicine-biology-chemistry separately.

6.1. *Further analyses*

6.1.1. *Exploring heterogeneity across fields*

We leverage on our large data sample of students representing all the STEM fields to explore cross-field heterogeneity. Table 4 shows some structural differences across fields. On average, students in Mathematics are the least productive, with 1.12 papers published during the training period, 0.88 average yearly citations received, and a network composed of 2.59 distinct co-authors. On the contrary, Ph.D. students enrolled in Medicine-biology-chemistry are the most productive. They show an average productivity of 3.22 publications, 2.96 yearly citations received, and a large network of 13.39 co-authors. Table B1, in Appendix B, reports the descriptive statistics of the complete set of explanatory variables by field.

Table 5 reports the estimations of the coefficients of Eq. (1) by field. Looking at the supervisors' biographic characteristics, differently from the regressions presented in Table 3, the relationship between supervisor's seniority and student's productivity is not statistically significant in Engineering and Physics. The supervisor's mentorship experience shows the same negative association with all the student's outcomes across fields: the greater the number of students previously mentored by the supervisor, the lower the student's productivity outcomes. Having a female supervisor relates positively to students' productivity in Engineering, while the effect is limited in the other fields. Specifically, having a female supervisor in Engineering is associated with 0.25 additional publications, 0.29 additional yearly citations received, and 0.79 additional co-authors. This result is particularly interesting due to the specificities of Engineering if compared with other disciplines. Indeed, female supervisors in engineering are rarer (only 14% of the supervisors are female scientists) than in other disciplines (17% in Mathematics, 28% in Medicine-biology-chemistry, and 17% in Physics) (Hunt, 2010). Moreover, the few female supervisors observed in Engineering, if compared to their male counterparts, are more productive than female supervisors in other disciplines.[26] We interpret these facts as the result of a selection process that leads only women with outstanding scientific competencies to overcome all the obstacles to reach a professorship position in a male-dominated discipline such as Engineering. These female supervisors' outstanding competencies are beneficial for the supervised students who show higher productivity.

When looking at the supervisors' academic characteristics, having a strong publication profile has a positive relationship with all the Ph.D. students' productivity outcomes across fields. The only exception is the negative relationship between the supervisor's number of publications and the student's network size in Mathematics, Medicine-biology-chemistry, and Physics. The number of citations received by the supervisor's publications has a positive relationship with all the student's productivity outcomes across fields. When we consider the supervisor's scientific network, the correlation between the supervisor's number of co-authors and all the Ph.D. student's productivity outcomes is positive in Medicine-biology-chemistry, while it is limited to the student's network in the other fields.

Results in Table 5 show that being mentored by a supervisor who benefited from an ANR grant is positively associated with all the Ph.D. students' productivity outcomes in Engineering and Physics. When we consider European grants, we find that they are positively associated with students' citations in Physics and Medicine-biology-chemistry. This latter result might be explained by the high student visibility gained in

[26] Looking at the publication score of female supervisors in engineering at the time of the students' enrollment, we find that their publication productivity is 77% of their male counterparts. In Mathematics is 69%, in Medicine-biology-chemistry is 65%, and in Physics is 69%.

Table 5
Regression results, by field. OLS estimates.

	Engineering			Mathematics			Medicine-biology-chemistry			Physics		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors
Supervisor characteristics												
Female supervisor	0.25***	0.29***	0.79***	−0.098**	−0.028	−0.13	−0.050	0.033	0.11	−0.028	0.048	0.73**
	(0.049)	(0.059)	(0.20)	(0.048)	(0.069)	(0.21)	(0.039)	(0.046)	(0.20)	(0.065)	(0.072)	(0.37)
Supervisor's seniority	0.010	−0.00093	0.046	0.029***	0.012	0.036	0.027**	−0.038**	0.14**	0.016	0.0094	0.051
	(0.013)	(0.015)	(0.052)	(0.011)	(0.016)	(0.049)	(0.014)	(0.016)	(0.069)	(0.016)	(0.018)	(0.090)
Supervisor's seniority^2	−0.00024	−0.00029	0.00042	−0.00100*	−0.00040	0.00028	−0.0022***	0.00030	−0.011***	−0.00084	−0.00073	−0.0060
	(0.00060)	(0.00071)	(0.0025)	(0.00057)	(0.00081)	(0.0025)	(0.00063)	(0.00074)	(0.0032)	(0.00078)	(0.00085)	(0.0043)
Mentorship experience	−0.013***	−0.011***	−0.033***	−0.0024	−0.010**	−0.0067	−0.035***	−0.0070*	−0.11***	−0.035***	−0.021***	−0.084**
	(0.0028)	(0.0034)	(0.012)	(0.0031)	(0.0045)	(0.014)	(0.0034)	(0.0040)	(0.017)	(0.0062)	(0.0068)	(0.035)
Supervisor's publications	0.034***	0.024***	0.039***	0.033***	0.012**	−0.069***	0.023***	0.0021	−0.12***	0.041***	0.024***	−0.040***
	(0.0026)	(0.0031)	(0.011)	(0.0033)	(0.0047)	(0.014)	(0.0020)	(0.0024)	(0.010)	(0.0026)	(0.0029)	(0.015)
Supervisor's average citations	0.019**	0.12***	0.018	0.020***	0.083***	0.078***	0.024***	0.27***	0.27***	0.055***	0.22***	0.21***
	(0.0077)	(0.0093)	(0.032)	(0.0051)	(0.0072)	(0.022)	(0.0061)	(0.0072)	(0.031)	(0.0092)	(0.010)	(0.052)
Supervisor's co-authors	−0.0044***	−0.0042***	0.015***	−0.0026**	0.00041	0.064***	0.0067***	0.0027***	0.11***	−0.0038***	−0.0024***	0.063***
	(0.00086)	(0.0010)	(0.0036)	(0.0010)	(0.0015)	(0.0044)	(0.00054)	(0.00064)	(0.0028)	(0.00064)	(0.00071)	(0.0036)
ANR grant	0.26***	0.42***	0.78**	0.14	0.10	1.35***	−0.16**	0.60***	−0.84**	0.53***	0.49***	2.37***
	(0.083)	(0.100)	(0.35)	(0.090)	(0.13)	(0.39)	(0.065)	(0.076)	(0.33)	(0.11)	(0.12)	(0.61)
EU grant	−0.012	−0.040	0.18	−0.35**	−0.21	−2.01***	−0.38***	0.33***	−1.46***	0.20	0.57***	−1.07
	(0.12)	(0.15)	(0.52)	(0.15)	(0.21)	(0.64)	(0.10)	(0.12)	(0.53)	(0.14)	(0.15)	(0.78)
Team characteristics												
With peers	0.12	−0.060	0.093	−0.049	0.085	−0.32	0.16**	0.31***	0.33	0.35***	0.34***	1.14**
	(0.081)	(0.097)	(0.34)	(0.077)	(0.11)	(0.33)	(0.067)	(0.079)	(0.34)	(0.093)	(0.10)	(0.52)
N. peers	−0.071***	−0.014	−0.22***	−0.048***	0.027*	−0.044	−0.27***	−0.13***	−1.01***	−0.14***	−0.048**	−0.53***
	(0.0100)	(0.012)	(0.041)	(0.010)	(0.015)	(0.044)	(0.015)	(0.018)	(0.077)	(0.021)	(0.023)	(0.12)
At least one female peer	0.093**	0.051	0.49***	0.032	−0.060	0.32*	−0.033	0.17***	0.25	−0.17***	−0.042	−0.022
	(0.039)	(0.047)	(0.16)	(0.043)	(0.061)	(0.18)	(0.046)	(0.054)	(0.23)	(0.061)	(0.067)	(0.34)
Average peers' seniority	−0.087***	−0.012	−0.16	−0.027	−0.045	0.030	−0.10***	−0.14***	−0.61***	−0.22***	−0.19***	−1.02***
	(0.031)	(0.037)	(0.13)	(0.031)	(0.044)	(0.13)	(0.029)	(0.034)	(0.15)	(0.042)	(0.046)	(0.23)
Peers' publications	0.12***	0.023	0.094	0.027	−0.079**	−0.47***	0.17***	−0.26***	−0.91***	0.23***	−0.051	−0.34
	(0.022)	(0.026)	(0.090)	(0.026)	(0.038)	(0.11)	(0.024)	(0.028)	(0.12)	(0.038)	(0.041)	(0.21)
Peers' average citations	−0.0048	0.018***	−0.028**	0.0012	0.0017	0.0027	0.011***	0.077***	0.11***	0.017***	0.075***	−0.0053
	(0.0033)	(0.0040)	(0.014)	(0.0038)	(0.0055)	(0.017)	(0.0032)	(0.0038)	(0.016)	(0.0059)	(0.0064)	(0.033)
Peers' co-authors	−0.0028	−0.0098**	0.055***	0.0092*	0.019***	0.14***	0.0085**	0.0088**	0.30***	−0.020***	−0.016**	0.16***
	(0.0036)	(0.0043)	(0.015)	(0.0048)	(0.0069)	(0.021)	(0.0038)	(0.0044)	(0.019)	(0.0058)	(0.0063)	(0.032)
Other controls												
French Top-20	−0.12**	0.050	−0.60***	−0.065	0.023	−0.21	−0.082**	0.052	−0.36*	0.18**	0.061	−0.25
	(0.050)	(0.060)	(0.21)	(0.047)	(0.067)	(0.20)	(0.037)	(0.043)	(0.19)	(0.079)	(0.086)	(0.44)
	0.035	0.00029	0.11	0.042**	0.064**	0.15**	−0.035*	0.017	−0.084	0.0044	0.021*	0.064

(*continued on next page*)

Table 5 (*continued*)

	Engineering			Mathematics			Medicine-biology-chemistry			Physics		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors	Publications	Average citations	Co-authors
Citation-weighted publications per affiliate												
	(0.027)	(0.033)	(0.11)	(0.018)	(0.025)	(0.076)	(0.018)	(0.022)	(0.093)	(0.010)	(0.011)	(0.057)
IDEX	−0.12**	−0.00096	−0.61**	0.082	−0.063	−0.0055	−0.092	−0.033	0.25	0.13	0.18*	0.47
	(0.059)	(0.070)	(0.24)	(0.063)	(0.090)	(0.27)	(0.067)	(0.078)	(0.34)	(0.087)	(0.096)	(0.49)
Department size [100 affiliates]	−0.0016	0.014***	−0.024*	0.00086	0.027***	−0.0051	0.0049***	−0.00011	0.022***	0.0060***	−0.0031	0.040***
	(0.0033)	(0.0040)	(0.014)	(0.0054)	(0.0077)	(0.023)	(0.0011)	(0.0012)	(0.0054)	(0.0017)	(0.0019)	(0.0096)
N. of Ph.D. students in the program	0.00013***	0.000022	0.00066***	0.00013***	0.000082**	0.00030***	−0.000076***	0.00028***	−0.00022	0.00019***	0.00038***	0.0011***
	(0.000028)	(0.000033)	(0.00012)	(0.000026)	(0.000036)	(0.00011)	(0.000029)	(0.000034)	(0.00015)	(0.000043)	(0.000047)	(0.00024)
Female student	−0.33***	−0.15***	−0.75***	−0.29***	−0.17***	−0.44**	−0.84***	−0.21***	−2.63***	−0.64***	−0.22***	−1.83***
	(0.039)	(0.047)	(0.16)	(0.040)	(0.058)	(0.17)	(0.034)	(0.040)	(0.17)	(0.052)	(0.057)	(0.29)
Co-supervision	0.073**	0.094**	0.21	0.035	0.12**	0.11	−0.23***	−0.22***	−1.74***	0.11**	0.14**	0.32
	(0.037)	(0.044)	(0.15)	(0.040)	(0.058)	(0.17)	(0.041)	(0.048)	(0.21)	(0.054)	(0.060)	(0.30)
Entry year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.85***	0.26*	1.89***	0.97***	0.12	1.52***	3.73***	1.87***	15.3***	1.76***	0.52**	6.22***
	(0.13)	(0.15)	(0.52)	(0.12)	(0.16)	(0.50)	(0.22)	(0.26)	(1.11)	(0.21)	(0.23)	(1.19)
Observations	16,519	16,519	16,519	11,450	11,450	11,450	35,038	35,038	35,038	14,136	14,136	14,136
R-squared	0.042	0.038	0.032	0.045	0.029	0.052	0.087	0.101	0.142	0.087	0.110	0.079

Note: Significance levels at ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are reported in parentheses. In an additional exercise, we calculate the p-values applying a multiple-inference adjustment to correct possible erroneous inferences due to the high number of hypotheses tested. We rely on the Romano-Wolf multiple hypothesis correction (Romano and Wolf, 2005a,b, 2016). The statistical significance of the coefficients remains almost unchanged (at standard significance levels) across disciplines. The only notable exception is the coefficient of the variable number of *Peers' co-authors*, which loses its statistical significance in several regressions explaining student's productivity in Engineering, Mathematics, and Medicine-biology-chemistry. The exercise is available upon request.

these fields due to the collaboration with research teams in other European countries promoted by the international nature of European grants.

In all fields, the increase in the number of peers is associated with decreased student's productivity, with the only exception of the increase in citations received in Mathematics. Peers' seniority is associated with a productivity decrease of the focal student in Medicine-biology-chemistry and Physics, while it shows no correlation with productivity in Mathematics and a slightly negative correlation in Engineering. Having at least one female peer is associated with scattered productivity benefits across disciplines, except for Physics. In Engineering, having a female peer relates to an increase in the publication score and network size, in mathematics with an increase in the network size, in Medicine-biology-chemistry with an increase in the citations received.

Peers' academic characteristics show mixed effects on students' productivity outcomes. Interestingly, the peers' network size is particularly favorable for the student's productivity in Mathematics and Medicine-biology-chemistry, while the peers' average citations benefit the student's productivity in Medicine-biology-chemistry and Physics. The peers' publication productivity is positively associated with the focal student's publication productivity in Engineering, Medicine-biology-chemistry, and Physics.

Concerning the control variables, consistently with Waldinger's study (2010) on mathematicians, we show a positive influence of the department's prestige on Ph.D. students' productivity in Mathematics. However, we show also that this result does not hold for students in Engineering and Medicine-biology-chemistry. This finding highlights the importance of covering multiple fields when assessing the determinants of students' productivity.

6.1.2. Considering different types of publication outcomes

We construct our students' productivity measures considering all the publications produced by the student during her training period. However, those publications might result from different research activities. In particular, some publications might result from the joint work between the student and her supervisor, while others from autonomous work or form collaborations with other scientists. Furthermore, some publications might result from the core thesis work, while others might result from other research lines unrelated to the thesis.

This section presents two robustness checks to investigate how the environmental factors relate to these different types of students' publications. First, we select only publications listing among the authors, both the student's and her supervisor's name. Doing so, we identify the publications that result from the close collaboration between the student and her supervisor. Second, we isolate the publications deriving from the student's thesis work. To do that, we use a text analysis algorithm to compare the thesis and publications' content and select only the student's publications with similar content to her thesis manuscript.

On average, students co-author with their supervisors 1.76 publications, which corresponds to 74% of the publications attributed to the students in our main analysis. Student-supervisor coauthored publications receive on average 1.97 yearly citations and list 6.92 co-authors (see Table C1 of Appendix C). Re-estimating in Table C2 the models presented in Table 3 considering only student-supervisor coauthored publications, we find results consistent with Table 3 with a few exceptions. For instance, from Table C2, we observe that the coefficient of the variable *ANR grant* turns positive in the regression explaining publication quantity and student's coauthors. Indeed, having a supervisor awarded an ANR grant is associated with 0.16 additional student-supervisor publications and 0.71 additional co-authors.

In Appendix D, Table D1 reports the descriptive statistics of the three dependent variables calculated attributing to the student only publications similar to the thesis manuscript. To measure the similarity between the publications authored by the student and the content of her thesis manuscript, we rely on a text analysis algorithm[27] that compares publication abstracts with thesis abstracts (Mikolov et al., 2013). On average, students have 1.38 publications similar to their thesis manuscript, which corresponds to 58% of the publications attributed on average to the students in our main analysis. These publications receive 1.41 yearly citations and list 5.37 co-authors. Table D2 reports the regression estimates of Eq. (1) using the three dependent variables considering only publications similar to the thesis. The regression results are consistent with our main analysis reported in Table 3, with a few exceptions. Like when looking only at the co-authored publications with the supervisor, also in the case of publications similar to the thesis, having a supervisor who is the principal investigator of an ANR grant positively correlates with student's publication quantity and co-authors. We interpret these results as the consequence of the pressure to publish experienced by ANR recipients who have to deliver publication outcomes as results of their project funded by the ANR agency. Therefore, supervisors with ANR grants tend to involve students in their projects, asking them to develop a thesis on ANR project topics and co-authoring with them. This involvement leads students to have a higher number of publications similar to the thesis and co-authored with the supervisor.

Interestingly, we observe also that the relationship between the supervisor's seniority and the Ph.D. student's productivity turns into a U-shaped relationship when we limit the analysis to publications similar to the thesis topics. This result differs from the inverted U-shaped relationship observed in the main analysis in Table 3. The change of the relationship between supervisor's seniority and the Ph.D. student's productivity might result from the evolution of the mentorship style along the supervisor's career. Specifically, mid-career supervisors who intend to boost their productivity under the pressure of being promoted from associate to full professors might consider students as lab workforce involving them in several projects, even not directly linked to their thesis, and co-author publications with them (Mangematin and Robin, 2003; Shibayama, 2019). As a result, when looking at the overall number of students' publications (the analysis reported in Table 3), and at the number of publications co-authored with the supervisors (the analysis reported in Table C2, in Appendix C), we observe an inverted U-shaped relationship between the supervisor's seniority and the student's productivity. On the contrary, this relationship turns in a U-shaped form when we focus on students' publications similar to the thesis content (Table D2, in Appendix D). Indeed, students considered as lab workforce by mid-career supervisors might lower the number of articles related to their thesis subject in favor of articles related to the supervisors' projects.

7. Conclusion

Students, directors of Ph.D. programs, and policymakers urge to identify the environmental characteristics correlated to Ph.D. students' productivity. From the students' perspective, showing a high-quality publication record and having a well-established scientific network is essential to be competitive in the job market after graduation. At the same time, directors of Ph.D. programs and policymakers need to optimize the use of resources and guarantee effective training programs.

In this paper, we study how social environment characteristics influence the Ph.D. students' productivity during their training period using a dataset that covers the entire population of 77,143 Ph.D. students who graduated from French universities in STEM disciplines between 2000 and 2014.

We consider the supervisor and peers' biographic and academic characteristics as relevant social environment characteristics. Then, we measure the student's productivity by counting the number of articles published during the training period (publication quantity), calculating the average number of citations received by the published articles

[27] Appendix D provides details on the text analysis algorithm.

(publication quality), and counting the number of distinct co-authors during the training period (scientific network size).

Not surprisingly, we find that students in productive environments are more productive, according to almost all the productivity measures considered. Having a female supervisor is associated with higher student productivity in engineering, the most male-centered discipline in our sample. Surprisingly, mentorship experience is associated with lower Ph.D. student's productivity, while having a mid-career supervisor is associated with higher student productivity. Having a supervisor with a French or European research grant is associated with a higher number of citations received by the student. Sharing the training experience with large groups of peers penalizes student's productivity, most likely due to a decline in the quality of the mentorship activity in large groups. On the contrary, having freshman peers, peers who publish high-quality articles, and at least one female peer is positively associated with student's productivity.

Some of our results align with a recent survey conducted in France in 2021 involving more than eleven thousand Ph.D. students from all fields (Pommier et al., 2022). The survey aimed to explore the Ph.D. students' perception of French Ph.D. programs. Results show that most respondents favor small-size teamwork with 2–3 students per supervisor. Consistently, half of the students declaring a lack of thesis progress is mentored by supervisors having more than four Ph.D. students simultaneously. Furthermore, the report shows that students evaluate the supervisor's role as fundamental in supporting the thesis progress and ensuring the financial conditions to carry out the research work.

A caveat applies to our analysis, as to a large part of the existing literature on Ph.D. students' productivity. Our econometric approach does not strictly allow a causal interpretation of the relationships between dependent and independent variables in our regression exercises. Nonetheless, we believe that limited biases affect our estimates for three reasons. First, we reduce the omitted variable problem by including proxies for all the factors that the extant literature considers relevant in affecting students' productivity, such as supervisor, peers, department, and student's time-variant and time-invariant characteristics. Second, theory suggests that information asymmetry in student selection makes it unlikely to observe a correlation between students' unobserved intrinsic ability and supervisors' quality, which might generate a potential endogeneity issue (Mangematin, 2000). In line with the theory, the empirical literature shows a weak correlation between proxies for the student's ability and Ph.D. productivity (Belavy et al., 2020). Third, to further investigate the potential endogeneity issue, we included in our model specification a proxy for students' ability using data on the participation of the students in selective contests during high school. Including this variable does not affect our main results, confirming the low likelihood of biased estimates in our regression exercises.

Our results speak to Ph.D. students, directors of Ph.D. programs, and policymakers. On the one hand, our paper provides hints to the students who want to leverage the environmental factors to boost their productivity. On the other hand, our results provide directors of Ph.D. programs and policymakers with a framework to understand the determinants of effective training programs and find levers for designing policies that maximize students' productivity. Along these lines, our work can be exploited to design better Ph.D. programs. Using our regression estimates, we can simulate how the students' productivity varies according to environment characteristics' changes. For example, by increasing the supervisor's publications by one standard deviation, decreasing the number of peers by one student, and reducing the average experience of the supervisors by one standard deviation, we obtain that the student's predicted productivity increases by one publication, one citation, and four additional co-authors. According to these predictions and causally interpreting our regression results, we may suggest three policy interventions that can be applied in the short run to increase the effectiveness of the French Ph.D. training system. First, professors' requirements to access students' supervision might be revised. In France, professors who supervise Ph.D. students must obtain a habilitation, *Habilitation a Diriger des Recherches*. The habilitation is awarded mainly by looking at the professor's scientific achievements. Raising the threshold for obtaining the habilitation would ensure supervisors with a higher number of publications and, according to our results, more productive students.[28] Second, a rule limiting the number of supervised students might be introduced, reducing the average number of peers. Finally, scientists who have never mentored Ph.D. students and fulfilling the requirements to supervise should be incentivized to start the mentorship activity, reducing the overall average experience of the supervisors. Combining these three policy interventions would enhance the effectiveness of the current Ph.D. training programs.

CRediT authorship contribution statement

Authors equally contributed to the paper and are listed in alphabetical order.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank participants in the 1st GREDEG Ph.D. Workshop, 2020, Nice (France); 8th WICK Ph.D. Workshop, 2020, Turin (Italy); Eu-SPRI Early Career Conference, 2020, Valencia (Spain); 18th ACDD Augustin Cournot Doctoral Days, 2021, Strasbourg (France); 9th Skema KTO Paper Development Workshop, 2021, Nice (France); NBER conference, Investments in Early Career Scientists: Data and Research Gaps, November 5, 2021 (Boston). We thank Thierry Burger-Helmchen, Adrián Arias Diaz-Faes, Maryann Feldman, Cornelia Lawson, Francesco Lissoni, and Sofia Patsali for their thoughtful comments. The French government has supported this work through the UCA^{JEDI} and EUR DS4H Investments in the Future projects managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-0001 and ANR-17-EURE-0004. This work also received the financial support of GIGA (ANR-19-CE26-0014).

Appendix A

This appendix reports the correlation matrix of the regressors included in Table 3. We find the highest correlation values between the variables *Supervisor's publications* and *Supervisor's co-authors* (0.78) and between *Peers' publications* and *Peers' co-authors* (0.89). In an alternative specification of the model estimated in Table 3, we excluded *Supervisor's* and *Peers' co-authors* from the model. Moreover, based on a Variance Inflation Factor (VIF) multicollinearity test, we excluded the variable *Citation-weighted publications per affiliate* that gives the highest VIF value (6.34). The model estimates excluding these three variables are consistent with those of Table 3 (Estimates without variables showing high correlation are available upon request).

[28] We assume that raising the threshold for the habitation does not create an undersupply of supervisors.

Table A1
Variable correlation matrix (N = 77,143).

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]
[1] Female supervisor	1.00																					
[2] Supervisor's seniority	−0.01	1.00																				
[3] Mentorship experience	−0.09	0.12	1.00																			
[4] Supervisor's publications	−0.11	0.32	0.19	1.00																		
[5] Supervisor avg. citations	0.03	0.24	−0.05	0.15	1.00																	
[6] Supervisor's co-authors	−0.05	0.33	0.07	0.78	0.24	1.00																
[7] ANR grant	0.02	0.17	−0.02	0.09	0.15	0.11	1.00															
[8] EU grant	−0.03	0.03	0.01	0.11	0.06	0.10	0.02	1.00														
[9] With peers	−0.06	0.09	0.17	0.11	0.02	0.04	0.06	0.03	1.00													
[10] N. peers	−0.09	0.04	0.50	0.17	−0.03	0.03	0.03	0.02	0.41	1.00												
[11] At least one female peer	0.00	0.09	0.19	0.14	0.05	0.09	0.06	0.03	0.52	0.44	1.00											
[12] Average peers' seniority	−0.08	0.13	0.26	0.14	0.01	0.05	0.05	0.04	0.77	0.56	0.52	1.00										
[13] Peers' publications	−0.04	0.10	0.20	0.27	0.05	0.17	0.04	0.03	0.23	0.45	0.22	0.32	1.00									
[14] Peers' average citations	−0.03	0.10	0.12	0.24	0.14	0.20	0.08	0.04	0.17	0.32	0.17	0.24	0.76	1.00								
[15] Peers' co-authors	−0.03	0.11	0.14	0.24	0.08	0.22	0.05	0.03	0.21	0.37	0.20	0.28	0.89	0.76	1.00							
[16] French Top-20	0.06	0.02	0.03	0.04	0.08	0.05	0.04	0.04	0.00	0.04	0.05	0.00	0.04	0.04	0.03	1.00						
[17] Citation-weighted publications per affiliate	0.09	0.40	−0.09	0.14	0.26	0.25	0.20	0.01	−0.04	−0.12	0.02	−0.05	0.04	0.08	0.07	0.06	1.00					
[18] IDEX	0.05	0.27	0.01	0.02	0.14	0.09	0.18	−0.02	0.01	0.01	0.03	0.03	0.04	0.06	0.05	0.14	0.44	1.00				
[19] Department size [100 aff.]	0.14	0.17	−0.06	0.18	0.21	0.25	0.10	0.04	−0.07	−0.11	0.05	−0.08	0.05	0.09	0.08	0.33	0.50	0.24	1.00			
[20] N. Ph.D. stud. in program	0.07	0.20	−0.02	0.11	0.19	0.15	0.11	0.05	0.02	0.04	0.05	0.03	0.08	0.10	0.08	0.33	0.38	0.29	0.47	1.00		
[21] Female student	0.09	0.04	−0.03	0.05	0.06	0.08	0.02	0.00	−0.04	−0.06	0.05	−0.04	−0.01	0.00	0.01	0.06	0.11	0.03	0.18	0.05	1.00	
[22] Co-supervision	0.00	0.13	0.00	−0.02	0.00	−0.02	0.04	−0.02	0.00	0.00	−0.01	0.03	−0.01	−0.01	−0.01	−0.12	0.08	0.06	−0.14	−0.04	−0.01	1.00

Appendix B

Table B1
Descriptive statistics of the explanatory variables, by field.

77,143 Ph.D. students	Engineering				Mathematics				Medicine-biology-chemistry				Physics			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Supervisor characteristics																
Female supervisor	0.14	0.35	0.00	1.00	0.17	0.37	0.00	1.00	0.28	0.45	0.00	1.00	0.17	0.38	0.00	1.00
Supervisor's seniority	11.11	5.07	0.00	21.00	9.89	5.46	0.00	21.00	12.20	4.99	0.00	21.00	11.47	5.53	0.00	21.00
Mentorship experience	4.41	7.44	0.00	114.00	3.97	7.48	0.00	114.00	2.37	5.49	0.00	184.00	2.56	4.71	0.00	108.0
Supervisor's publications	11.01	11.93	0.00	98.00	6.92	9.46	0.00	93.00	16.86	15.69	0.00	100.00	13.91	14.07	0.00	100.0
Supervisor's average citations	1.76	2.27	0.00	87.17	1.54	3.58	0.00	127.87	2.95	3.08	0.00	113.09	2.28	2.88	0.00	98.22
Supervisor's co-authors	22.72	34.31	0.00	498.00	13.08	29.38	0.00	468.00	50.82	56.15	0.00	499.0	40.36	54.95	0.00	498.00
ANR grant	0.05	0.21	0.00	1.00	0.04	0.20	0.00	1.00	0.08	0.28	0.00	1.00	0.06	0.23	0.00	1.00
EU grant	0.02	0.14	0.00	1.00	0.01	0.12	0.00	1.00	0.03	0.16	0.00	1.00	0.03	0.18	0.00	1.00
Team characteristics																
With peers	0.89	0.31	0.00	1.00	0.84	0.37	0.00	1.00	0.76	0.43	0.00	1.00	0.77	0.42	0.00	1.00
N. peers	2.54	2.48	0.00	28.25	2.27	2.73	0.00	28.25	1.33	1.68	0.00	28.25	1.49	1.80	0.00	30.00
At least one female peer	0.53	0.50	0.00	1.00	0.48	0.50	0.00	1.00	0.55	0.50	0.00	1.00	0.45	0.50	0.00	1.00
Average peers' seniority	1.91	0.91	0.00	3.48	1.76	1.00	0.00	3.43	1.46	1.07	0.00	3.44	1.51	1.06	0.00	3.56
Peers' publications	0.88	1.94	0.00	27.25	0.68	1.69	0.00	29.75	0.85	1.78	0.00	41.00	0.70	1.55	0.00	25.75
Peers' average citations	2.57	8.41	0.00	353.15	1.88	7.38	0.00	187.40	3.15	8.56	0.00	266.58	2.47	6.99	0.00	150.54
Peers' co-authors	4.21	10.87	0.00	190.75	3.18	9.60	0.00	176.25	4.77	10.54	0.00	187.25	3.66	9.29	0.00	150.00
Other controls																
French Top-20	0.24	0.43	0.00	1.00	0.53	0.50	0.00	1.00	0.49	0.50	0.00	1.00	0.19	0.39	0.00	1.00
Citation-weighted publications per affiliate	3.96	1.61	0.38	10.72	3.71	1.55	0.81	10.61	8.54	3.41	0.93	17.58	11.43	5.40	1.35	35.05
IDEX	0.15	0.36	0.00	1.00	0.19	0.39	0.00	1.00	0.19	0.39	0.00	1.00	0.19	0.40	0.00	1.00
Department size [100 affiliates]	9.54	6.12	0.04	27.99	6.57	4.55	0.10	21.54	49.33	32.74	0.18	114.46	20.87	18.64	0.15	64.30
N. of Ph.D. students in the program	753.04	680.93	5.00	2973.0	1000.73	795.82	1.00	2973.0	1138.96	803.44	1.00	2973.0	1173.13	840.62	1.00	2973.0
Female student	0.25	0.43	0.00	1.00	0.27	0.44	0.00	1.00	0.53	0.50	0.00	1.00	0.33	0.47	0.00	1.00
Co-supervision	0.39	0.49	0.00	1.00	0.31	0.46	0.00	1.00	0.26	0.44	0.00	1.00	0.35	0.48	0.00	1.00
Entry year	2005.20	4.13	1997.0	2011.0	2005.47	4.08	1997.0	2011.0	2004.93	4.21	1997.0	2011.0	2005.23	4.30	1997.00	2011.0
Observations	16,519				11,450				35,038				14,136			

Appendix C

This appendix reports a robustness check in which we only select the publications of the Ph.D. student co-authored with her supervisor to calculate our dependent variables. Using this selection criterion, we find that 59.79% of students have at least one paper co-authored with the supervisor during the training period.

Table C1 shows the descriptive statistics of the newly calculated dependent variables, while Table C2 shows the regression results.

Table C1
Descriptive statistics of the students' productivity outcomes. Publication attribution is based on the co-authorship with the supervisor.

Dependent variables 77,143 Ph.D. students	Mean	Sd	Min	Max
Publications	1.76	2.33	0.00	20.00
Average citations	1.97	3.59	0.00	170.42
Co-authors	6.92	12.27	0.00	195.00

Table C2
Regression results. Publication attribution is based on the co-authorship with the supervisor. OLS estimates.

	(1)	(2)	(3)
	Publications	Average citations	Co-authors
Supervisor characteristics			
Female supervisor	0.028	0.069**	0.37***
	(0.019)	(0.030)	(0.10)
Supervisor's seniority	0.11***	0.067***	0.33***
	(0.0055)	(0.0086)	(0.029)
Supervisor's seniority^2	−0.0048***	−0.0032***	−0.015***
	(0.00026)	(0.00041)	(0.0014)
Mentorship experience	−0.019***	−0.0081***	−0.041***
	(0.0015)	(0.0023)	(0.0076)
Supervisor's publications	0.027***	0.0078***	−0.083***
	(0.00094)	(0.0015)	(0.0049)
Supervisor's average citations	0.038***	0.21***	0.23***
	(0.0028)	(0.0043)	(0.014)
Supervisor's co-authors	0.00073***	0.0019***	0.074***
	(0.00026)	(0.00041)	(0.0013)
ANR grant	0.16***	0.58***	0.71***
	(0.033)	(0.051)	(0.17)
EU grant	−0.15***	0.27***	−0.94***
	(0.050)	(0.078)	(0.26)
Team characteristics			
With peers	0.20***	0.23***	0.53***
	(0.031)	(0.049)	(0.16)
N. peers	−0.077***	−0.046***	−0.28***
	(0.0054)	(0.0085)	(0.028)
At least one female peer	−0.026	0.071**	0.16
	(0.019)	(0.030)	(0.100)
Average peers' seniority	−0.16***	−0.14***	−0.64***
	(0.013)	(0.021)	(0.068)
Peers' publications	0.074***	−0.17***	−0.67***
	(0.011)	(0.017)	(0.055)
Peers' average citations	0.011***	0.059***	0.067***
	(0.0015)	(0.0024)	(0.0080)
Peers' co-authors	0.0011	0.0039	0.17***
	(0.0017)	(0.0027)	(0.0090)
Other controls			
French Top-20	−0.063***	0.044	−0.40***
	(0.018)	(0.028)	(0.093)
Citation-weighted publications per affiliate	0.013***	0.026***	0.13***
	(0.0043)	(0.0068)	(0.023)
IDEX	−0.021	0.0088	0.035
	(0.027)	(0.043)	(0.14)
Department size [100 affiliates]	−0.00042	0.0016**	0.0057**
	(0.00043)	(0.00068)	(0.0023)
N. of Ph.D. students in the program	0.000047***	0.00020***	0.00028***
	(0.000012)	(0.000019)	(0.000062)
Female student	−0.36***	−0.18***	−1.05***
	(0.016)	(0.026)	(0.084)
Co-supervision	−0.094***	−0.077***	−0.60***
	(0.018)	(0.028)	(0.091)
Engineering	0.35***	0.45***	0.92***
	(0.027)	(0.042)	(0.14)

(*continued on next page*)

Table C2 (*continued*)

	(1) Publications	(2) Average citations	(3) Co-authors
Physics	0.72***	0.61***	1.78***
	(0.042)	(0.066)	(0.22)
Medicine-biology-chemistry	1.41***	1.44***	5.24***
	(0.033)	(0.051)	(0.17)
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	0.25***	−0.25***	1.14***
	(0.056)	(0.088)	(0.29)
Observations	77,143	77,143	77,143
R-squared	0.172	0.135	0.193

Note: Significance levels at ***p < 0.01, **p < 0.05, *p < 0.1. Standard errors are reported in parentheses.

Appendix D

This appendix reports a robustness check in which we only select the Ph.D. student's publications showing high similarity with the abstract of the thesis manuscript. We expect that a large part of students' publications during the training period derives from the thesis research work.

To measure the similarity between the student's thesis and her publications, we rely on a text analysis algorithm (Mikolov et al., 2013). Specifically, we proceed in two steps. First, we use the Word2Vec algorithm to attribute to each word its vectorial representation according to the word's semantic meaning. To do so, we generate a vocabulary of 411,525 words retrieved from all the distinct words appearing in the abstracts of 1,284,753 STEM publications in English authored by French researchers in 1990–2018. Then, we use the co-occurrence of words in the articles' abstract to train our algorithm and provide for each word a 100-dimension vectorial representation (Rong, 2014). Each dimension in the vectorial space represents a latent dimension of the word's semantic meaning. Once we generate a vocabulary that allows us to translate words into vectors, we attribute to each word appearing within theses and student publications' abstracts its vectorial representation. Therefore, after this operation, theses and students' publications are represented by a series of vectors corresponding to words, each of which is a point in the 100-dimension vectorial space. In order to obtain the vectorial representation of the entire text documents, we calculate the centroid of all the vectors representing each document. When all the documents are represented by a unique vector, we calculate the cosine similarity between the vectors representing the student thesis and the vectors representing the student's publications. Cosine similarity values range from −1 (highly dissimilar documents) to +1 (highly similar documents). We consider a thesis similar to a publication if the cosine similarity value exceeds the threshold of 0.8. Once calculated the similarity between documents, we attribute to students only papers similar to her Ph.D. thesis. We end up with 44.27% of students having at least one paper attributed.

Table D1 shows the descriptive statistics of the newly calculated dependent variables, while Table D2 shows the regression results.

Table D1
Descriptive statistics of the students' productivity outcomes. Publication attribution is based on similarity between student's thesis and publications.

Dependent variables 77,143 Ph.D. students	Mean	Sd	Min	Max
Publications	1.38	2.30	0.00	20.00
Average citations	1.41	3.09	0.00	120.24
Co-authors	5.37	11.82	0.00	200.00

Table D2
Regression results. Publication attribution is based on similarity between student's thesis and publications. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor characteristics			
Female supervisor	0.035*	0.087***	0.33***
	(0.019)	(0.026)	(0.098)
Supervisor's seniority	−0.015***	−0.033***	−0.085***
	(0.0055)	(0.0075)	(0.028)
Supervisor's seniority^2	0.00075***	0.0012***	0.0042***
	(0.00026)	(0.00036)	(0.0013)
Mentorship experience	−0.011***	−0.0028	−0.019**
	(0.0015)	(0.0020)	(0.0075)
Supervisor's publications	0.016***	0.0015	−0.078***
	(0.00094)	(0.0013)	(0.0048)
Supervisor's average citations	0.025***	0.14***	0.17***
	(0.0028)	(0.0038)	(0.014)
Supervisor's co-authors	0.0013***	0.0016***	0.058***
	(0.00026)	(0.00036)	(0.0013)
ANR grant	0.19***	0.69***	0.99***
	(0.033)	(0.045)	(0.17)

(*continued on next page*)

Table D2 (*continued*)

	(1)	(2)	(3)
	Publications	Average citations	Co-authors
EU grant	−0.12**	0.15**	−0.85***
	(0.050)	(0.068)	(0.25)
Team characteristics			
With peers	0.16***	0.14***	0.53***
	(0.031)	(0.043)	(0.16)
N. peers	−0.072***	−0.035***	−0.24***
	(0.0054)	(0.0074)	(0.027)
At least one female peer	−0.012	0.056**	0.16
	(0.019)	(0.026)	(0.098)
Average peers' seniority	−0.092***	−0.063***	−0.49***
	(0.013)	(0.018)	(0.066)
Peers' publications	0.086***	−0.089***	−0.30***
	(0.011)	(0.015)	(0.054)
Peers' average citations	−0.00032	0.028***	0.0048
	(0.0015)	(0.0021)	(0.0078)
Peers' co-authors	0.0013	0.0038	0.12***
	(0.0017)	(0.0024)	(0.0089)
Other controls			
French Top-20	−0.24***	−0.17***	−1.00***
	(0.018)	(0.024)	(0.091)
Citation-weighted publications per affiliate	0.070***	0.073***	0.36***
	(0.0044)	(0.0060)	(0.022)
IDEX	0.050*	0.19***	0.59***
	(0.027)	(0.037)	(0.14)
Department size [100 affiliates]	−0.0032***	−0.0038***	−0.012***
	(0.00043)	(0.00059)	(0.0022)
N. of Ph.D. students in the program	−0.00017***	−0.000097***	−0.00054***
	(0.000012)	(0.000016)	(0.000061)
Female student	−0.32***	−0.15***	−0.96***
	(0.016)	(0.022)	(0.083)
Co-supervision	0.076***	0.044*	0.0055
	(0.018)	(0.024)	(0.089)
Engineering	0.19***	0.33***	0.79***
	(0.027)	(0.037)	(0.14)
Physics	0.14***	0.034	0.27
	(0.042)	(0.058)	(0.21)
Medicine-biology-chemistry	0.69***	0.85***	3.42***
	(0.033)	(0.045)	(0.17)
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	1.19***	0.72***	3.66***
	(0.056)	(0.077)	(0.29)
Observations	77,143	77,143	77,143
R-squared	0.146	0.114	0.165

Note: Significance levels at ***p < 0.01, **p < 0.05, *p < 0.1. Standard errors are reported in parentheses.

Appendix E

This appendix reports a regression exercise where we include a proxy for the student's intrinsic ability among the control variables. Specifically, we identify in our study sample the students who have participated in three well-known contests during the high school period: the International Mathematical Olympiad (IMO), *Les Olympiades Nationales de Mathématiques* (the national French Mathematical Olympiad), and *le Kangourou des mathématiques* (a French national mathematical contest).[29] These contests are organized both at the national and international level, and students who show particular abilities during their high school studies are selected to participate. We argue that this variable is a good proxy for students' intrinsic ability, interest, and motivation in schooling and education.

We found 138 Ph.D. students who participated in at least one of the three contests and were mentioned in the contests' final ranking (with or without winning a medal). In our econometric exercise, we identify those students with the dummy variable *Math Olympiad* that equals one if the student participated in at least one of the three contests, zero otherwise. As expected, we find that a large share of students ends up doing a Ph.D. in Mathematics (53%); nonetheless, a non-negligible share did a Ph.D. in engineering (19%), Physics (12%), and Medicine-biology-chemistry (16%).

Table E1 reports the regression exercise results, including the *Math Olympiad* dummy variable among the controls. The results concerning the supervisor's and peers' characteristics are in line with those presented in Table 3 in our main analysis, and the dummy *Math Olympiad* is never significant in all the three econometric models considered.

We conclude that including a proxy for the student's ability does not change the impact of the environmental characteristics on the student's scientific productivity. These results are consistent with previous literature findings (Aristizábal, 2021; Belavy et al., 2020; Mangematin, 2000).

[29] Data for the International Mathematical Olympiad (IMO) are available from 1981 to 2009, for *Les Olympiades Nationales de Mathématiques* from 2001 to 2007, and for *le Kangourou des mathématiques* from 2005 to 2007.

Table E1
Regression results. Including a proxy for the student's ability. OLS estimates.

	(1)	(2)	(3)
	Publications	Average citations	Co-authors
Student's ability			
Math Olympiad	0.19	−0.0094	−0.87
	(0.24)	(0.28)	(1.19)
Supervisor characteristics			
Female supervisor	−0.0049	0.074**	0.31**
	(0.025)	(0.030)	(0.13)
Supervisor's seniority	0.037***	0.0071	0.11***
	(0.0072)	(0.0085)	(0.036)
Supervisor's seniority^2	−0.0019***	−0.00096**	−0.0067***
	(0.00034)	(0.00040)	(0.0017)
Mentorship experience	−0.018***	−0.0072***	−0.037***
	(0.0019)	(0.0023)	(0.0097)
Supervisor's publications	0.027***	0.0070***	−0.10***
	(0.0012)	(0.0015)	(0.0062)
Supervisor's average citations	0.031***	0.20***	0.21***
	(0.0036)	(0.0043)	(0.018)
Supervisor's co-authors	0.0028***	0.0014***	0.091***
	(0.00034)	(0.00040)	(0.0017)
ANR grant	0.0050	0.54***	0.22
	(0.043)	(0.050)	(0.21)
EU grant	−0.19***	0.33***	−1.28***
	(0.065)	(0.077)	(0.33)
Team characteristics			
With peers	0.13***	0.24***	0.25
	(0.041)	(0.048)	(0.21)
N. peers	−0.12***	−0.042***	−0.39***
	(0.0071)	(0.0083)	(0.036)
At least one female peer	−0.028	0.073**	0.21*
	(0.025)	(0.030)	(0.13)
Average peers' seniority	−0.14***	−0.13***	−0.63***
	(0.017)	(0.020)	(0.086)
Peers' publications	0.13***	−0.15***	−0.64***
	(0.014)	(0.016)	(0.070)
Peers' average citations	0.0065***	0.056***	0.049***
	(0.0020)	(0.0024)	(0.010)
Peers' co-authors	0.0029	0.0017	0.21***
	(0.0023)	(0.0027)	(0.011)
Other controls			
French Top-20	−0.0084	0.068**	−0.36***
	(0.023)	(0.028)	(0.12)
Citation-weighted publications per affiliate	0.012**	0.026***	0.14***
	(0.0057)	(0.0067)	(0.029)
IDEX	−0.056	0.031	−0.031
	(0.036)	(0.042)	(0.18)
Department size [100 affiliates]	0.00081	0.0014**	0.013***
	(0.00057)	(0.00067)	(0.0029)
N. of Ph.D. students in the program	0.000092***	0.00023***	0.00038***
	(0.000016)	(0.000019)	(0.000079)
Female student	−0.64***	−0.19***	−1.84***
	(0.021)	(0.025)	(0.11)
Co-supervision	−0.065***	−0.042	−0.66***
	(0.023)	(0.027)	(0.12)
Engineering	0.18***	0.40***	0.99***
	(0.035)	(0.041)	(0.18)
Physics	0.77***	0.57***	2.45***
	(0.055)	(0.065)	(0.28)
Medicine-biology-chemistry	1.54***	1.39***	6.44***
	(0.043)	(0.050)	(0.22)
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	1.23***	0.23***	3.84***
	(0.074)	(0.087)	(0.37)
Observations	77,143	77,143	77,143
R-squared	0.140	0.128	0.174

Note: Significance levels at ***p < 0.01, **p < 0.05, *p < 0.1. Standard errors are reported in parentheses.

Appendix F

This appendix searches for non-linear associations between student outcomes, number of peers, and supervisor's mentorship experience.

Table F1 investigates the possible nonlinear association between the supervisor's mentorship experience and the student's productivity outcomes. We calculate eight dummy variables, each of which takes value one if the number of Ph.D. students successfully supervised before the focal student enrollment equals 1, 2, 3, 4, 5, 6, 7, or is larger than 7. Specifically, The variable *Mentorship experience* = *1*, takes value one if the supervisor has only one Ph.D. student who graduated before the enrollment of the focal student. The variable equals one for 15.1% of supervisors. Similarly, we calculate *Mentorship experience* = *2* (9.7%), *Mentorship experience* = *3* (7.0%), *Mentorship experience* = *4* (5.1%), *Mentorship experience* = *5* (3.9%), *Mentorship experience* = *6* (2.9%), *Mentorship experience* = *7* (2.4%), and *Mentorship experience* > *7* (11.7%). The reference case, represented by the vertical line centered in zero, is when the supervisor has *No mentorship experience* (42.1%). We find that results in Table F1 confirm a negative association between the supervisor's mentorship experience and the three student's outcomes, as shown in Table 3. Fig. 1 in the main text depicts the association between mentorship experience and Ph.D. students' productivity.

Table F2 investigates the possible nonlinear association between the student's number of peers and the student's productivity outcomes. Based on the values of the variables *N. peers*, we calculated six dummy variables. The first variable, $0 < N.\ peers \leq 1$, takes value one for all those students having between 0 (excluded) and 1 (included) peers during the Ph.D. period. The variable equals one for 29.9% of the students. Similarly, we calculate $1 < N.\ peers \leq 2$ (20.7% of the students), $2 < N.\ peers \leq 3$ (12.2%), $3 < N.\ peers \leq 4$ (6.9%), $4 < N.\ peers \leq 5$ (3.8%), and $N.\ peers > 5$ (6.4%). The reference case is when the focal student has *No peers* (20.0%). Table F2 shows a similar pattern as the one observed in the regressions in Table 3 in the main text. We find a positive association between $0 < N.\ peers \leq 1$ on the citations received by the student's articles. An increase in the number of peers is associated with a sharp decrease of the student's publications leading to −0.83 articles and −3.21 co-authors when the peer number exceeds 5 peers. Interestingly, a large number of peers are not associated with a significant decrease in citations.

We interpret the result on publication productivity as a loss of supervisor's attention to the student's work. In the case of many peers, the supervisor shares her limited time with many students reducing her support to each of them. A similar interpretation can apply to the citations received by the student's work. If the supervisor has up to 3 students at a time (the focal student + 2 peers) the quality of the student's work probably benefits from the supervisor's advice. Concerning the negative association between the number of peers and the number of co-authors, one possible explanation is that the student having many peers within the team has less incentive to look for other collaborators outside the team, reducing the probability of finding new co-authors or joining other research teams. Fig. 2 in the main text provides a visual representation of the association between the peer group size (*N. peers*) and the Ph.D. student's productivity.

Table F1
Regression with mentoring experience dummy variables.

Variables	(1)	(2)	(3)
	Publications	Average citations	Co-authors
Supervisor female	−0.010	0.071**	0.29**
	(0.025)	(0.030)	(0.13)
Supervisor's seniority	0.033***	0.0046	0.094***
	(0.0072)	(0.0085)	(0.036)
Supervisor's seniority^2	−0.0014***	−0.00073*	−0.0050***
	(0.00034)	(0.00040)	(0.0017)
Mentorship experience = 0	Ref.	Ref.	Ref.
Mentorship experience = 1	−0.074**	0.012	−0.0049
	(0.031)	(0.036)	(0.15)
Mentorship experience = 2	−0.23***	−0.100**	−0.48***
	(0.037)	(0.043)	(0.18)
Mentorship experience = 3	−0.26***	−0.070	−0.73***
	(0.042)	(0.050)	(0.21)
Mentorship experience = 4	−0.26***	−0.062	−0.65***
	(0.048)	(0.057)	(0.24)
Mentorship experience = 5	−0.38***	−0.18***	−1.14***
	(0.054)	(0.064)	(0.27)
Mentorship experience = 6	−0.32***	−0.14*	−1.16***
	(0.062)	(0.073)	(0.31)
Mentorship experience = 7	−0.52***	−0.19**	−1.76***
	(0.068)	(0.080)	(0.34)
Mentorship experience > 7	−0.55***	−0.25***	−1.44***
	(0.039)	(0.046)	(0.19)
Supervisors' publications	0.028***	0.0077***	−0.099***
	(0.0012)	(0.0015)	(0.0062)
Supervisors' average citations	0.030***	0.20***	0.21***
	(0.0036)	(0.0043)	(0.018)
Supervisors' co-authors	0.0026***	0.0013***	0.090***
	(0.00034)	(0.00040)	(0.0017)
ANR grant	0.0013	0.53***	0.20
	(0.043)	(0.050)	(0.21)
EU grant	−0.17***	0.34***	−1.23***
	(0.065)	(0.077)	(0.33)
With peers	0.11***	0.23***	0.19
	(0.041)	(0.048)	(0.21)
N. peers	−0.12***	−0.040***	−0.37***
	(0.0068)	(0.0080)	(0.034)

(*continued on next page*)

Table F1 (*continued*)

Variables	(1) Publications	(2) Average citations	(3) Co-authors
At least one female peer	−0.015	0.079***	0.25**
	(0.025)	(0.030)	(0.13)
Average peers' seniority	−0.10***	−0.12***	−0.53***
	(0.017)	(0.020)	(0.087)
Peers' publications	0.13***	−0.15***	−0.64***
	(0.014)	(0.016)	(0.070)
Peers' average citations	0.0067***	0.056***	0.049***
	(0.0020)	(0.0024)	(0.010)
Peers' co-authors	0.0034	0.0019	0.21***
	(0.0023)	(0.0027)	(0.011)
French Top-20	−0.0080	0.069**	−0.36***
	(0.023)	(0.028)	(0.12)
Citation-weighted publications per affiliate	0.011*	0.026***	0.14***
	(0.0057)	(0.0067)	(0.029)
IDEX	−0.039	0.038	0.023
	(0.036)	(0.042)	(0.18)
Department size [100 affiliates]	0.0011*	0.0015**	0.014***
	(0.00057)	(0.00067)	(0.0029)
N. of Ph.D. students in the program	0.000074***	0.00022***	0.00033***
	(0.000016)	(0.000019)	(0.000079)
Female student	−0.64***	−0.19***	−1.84***
	(0.021)	(0.025)	(0.11)
Co-supervision	−0.066***	−0.041	−0.66***
	(0.023)	(0.027)	(0.12)
Engineering	0.17***	0.39***	0.96***
	(0.035)	(0.041)	(0.18)
Physics	0.76***	0.56***	2.41***
	(0.055)	(0.065)	(0.28)
Medicine-biology-chemistry	1.50***	1.37***	6.32***
	(0.043)	(0.051)	(0.22)
Mathematics	Ref.	Ref.	Ref.
Constant	1.54***	−0.12	2.51***
	(0.068)	(0.080)	(0.34)
Observations	77,143	77,143	77,143
R-squared	0.142	0.128	0.175
Entry year dummies	Yes	Yes	Yes

Note: Significance levels at ***p < 0.01, **p < 0.05, *p < 0.1. Standard errors are reported in parentheses.

Table F2
Regression with dummy variables for the peer group size.

Variables	(1) Publications	(2) Average citations	(3) Coauthors
Supervisor female	−0.0073	0.074**	0.29**
	(0.025)	(0.030)	(0.13)
Supervisor's seniority	0.036***	0.0068	0.11***
	(0.0072)	(0.0085)	(0.036)
Supervisor's seniority^2	−0.0019***	−0.00094**	−0.0066***
	(0.00034)	(0.00040)	(0.0017)
Mentorship experience	−0.022***	−0.0086***	−0.052***
	(0.0018)	(0.0022)	(0.0093)
Supervisors' publications	0.027***	0.0072***	−0.10***
	(0.0012)	(0.0015)	(0.0062)
Supervisors' average citations	0.031***	0.20***	0.21***
	(0.0036)	(0.0043)	(0.018)
Supervisors' co-authors	0.0027***	0.0014***	0.090***
	(0.00034)	(0.00040)	(0.0017)
ANR grant	0.0039	0.54***	0.23
	(0.043)	(0.050)	(0.21)
EU grant	−0.18***	0.33***	−1.22***
	(0.065)	(0.077)	(0.33)
No peers	Ref.	Ref.	Ref.
$0 < \text{N. peers} \leq 1$	0.048	0.21***	−0.12
	(0.042)	(0.049)	(0.21)
$1 < \text{N. peers} \leq 2$	−0.065	0.22***	−0.79***
	(0.057)	(0.067)	(0.28)
$2 < \text{N. peers} \leq 3$	−0.27***	0.12	−1.65***
	(0.063)	(0.074)	(0.32)
$3 < \text{N. peers} \leq 4$	−0.41***	0.032	−2.13***
	(0.069)	(0.081)	(0.35)

(continued on next page)

Table F2 (*continued*)

Variables	(1) Publications	(2) Average citations	(3) Coauthors
$4 < \text{N. peers} \leq 5$	−0.58***	−0.060	−2.58***
	(0.078)	(0.092)	(0.39)
N. peers > 5	−0.83***	−0.060	−3.21***
	(0.074)	(0.088)	(0.37)
At least one female peer	−0.0045	0.082***	0.37***
	(0.026)	(0.030)	(0.13)
Average peers' seniority	−0.12***	−0.13***	−0.45***
	(0.019)	(0.023)	(0.098)
Peers' publications	0.12***	−0.15***	−0.65***
	(0.014)	(0.016)	(0.069)
Peers' average citations	0.0070***	0.056***	0.051***
	(0.0020)	(0.0024)	(0.010)
Peers' co-authors	0.0035	0.0018	0.21***
	(0.0023)	(0.0027)	(0.011)
French Top-20	−0.014	0.067**	−0.38***
	(0.023)	(0.028)	(0.12)
Citation-weighted publications per affiliate	0.012**	0.026***	0.14***
	(0.0057)	(0.0067)	(0.029)
IDEX	−0.055	0.030	−0.032
	(0.036)	(0.042)	(0.18)
Department size [100 affiliates]	0.00091	0.0014**	0.013***
	(0.00057)	(0.00067)	(0.0029)
N. of Ph.D. students in the program	0.000084***	0.00022***	0.00036***
	(0.000016)	(0.000019)	(0.000079)
Female student	−0.64***	−0.19***	−1.85***
	(0.021)	(0.025)	(0.11)
Co-supervision	−0.062***	−0.040	−0.65***
	(0.023)	(0.027)	(0.12)
Engineering	0.18***	0.40***	1.04***
	(0.035)	(0.041)	(0.18)
Physics	0.77***	0.57***	2.42***
	(0.055)	(0.065)	(0.28)
Medicine-biology-chemistry	1.52***	1.38***	6.36***
	(0.043)	(0.051)	(0.22)
Mathematics	Ref.	Ref.	Ref.
Constant	1.44***	−0.16**	2.30***
	(0.067)	(0.079)	(0.34)
Observations	77,143	77,143	77,143
R-squared	0.141	0.128	0.175
Entry year dummies	Yes	Yes	Yes

Note: Significance levels at ***p < 0.01, **p < 0.05, *p < 0.1. Standard errors are reported in parentheses.

References

Agarwal, R., Gaule, P., 2020. Invisible geniuses: could the knowledge frontier advance faster? Am. Econ. Rev. Insights 2, 409–424. https://doi.org/10.1257/aeri.20190457.

Alberts, B., Kirschner, M.W., Tilghman, S., Varmus, H., 2014. Rescuing US biomedical research from its systemic flaws. PNAS 111, 5773–5777. https://doi.org/10.1073/pnas.1404402111.

Allison, P.D., Stewart, J.A., 1974. Productivity differences among scientists: evidence for accumulative advantage. Am. Sociol. Rev. 39, 596–606.

Aristizábal, N., 2021. I bombed the GRE—but I'm thriving as a Ph.D. student [WWW Document]. Science. URL. AAAS (accessed 6.17.21). https://www.sciencemag.org/careers/2021/06/i-bombed-gre-i-m-thriving-phd-student.

Ayoubi, C., Pezzoni, M., Visentin, F., 2017. At the origins of learning: absorbing knowledge flows from within the team. J. Econ. Behav. Organ. 134, 374–387.

Azoulay, P., Zivin, J.S.G., Wang, J., 2010. Superstar extinction. Q. J. Econ. 42.

Bandiera, O., Barankay, I., Rasul, I., 2009. Social connections and incentives in the workplace: evidence from personnel data. Econometrica 77, 1047–1094.

Baruffaldi, S., Visentin, F., Conti, A., 2016. The productivity of science & engineering PhD students hired from supervisors' networks. Res. Policy 45, 785–796. https://doi.org/10.1016/j.respol.2015.12.006.

Belavy, D.L., Owen, P.J., Livingston, P.M., 2020. Do successful PhD outcomes reflect the research environment rather than academic ability? PLOS ONE 15, e0236327. https://doi.org/10.1371/journal.pone.0236327.

Benedictus, R., Miedema, F., Ferguson, M.W.J., 2016. Fewer numbers, better science. Nature 538, 453–455. https://doi.org/10.1038/538453a.

Bercovitz, J., Feldman, M., 2011. The mechanisms of collaboration in inventive teams: composition, social networks, and geography. Res. Policy 40, 81–93. https://doi.org/10.1016/j.respol.2010.09.008.

Black, G.C., Stephan, P.E., 2010. The economics of university science and the role of foreign graduate students and postdoctoral scholars. In: American Universities in a Global Market. University of Chicago Press, pp. 129–161.

Bozeman, B., Corley, E., 2004. Scientists' collaboration strategies: implications for scientific and technical human capital. Res. Policy 33, 599–616. https://doi.org/10.1016/j.respol.2004.01.008.

Brischoux, F., Angelier, F., 2015. Academia's never-ending selection for productivity. Scientometrics 103, 333–336. https://doi.org/10.1007/s11192-015-1534-5.

Broström, A., 2019. Academic breeding grounds: home department conditions and early career performance of academic researchers. Res. Policy 48, 1647–1665. https://doi.org/10.1016/j.respol.2019.03.009.

Carayol, N., Matt, M., 2004. Does research organization influence academic production? Res. Policy 33, 1081–1102. https://doi.org/10.1016/j.respol.2004.03.004.

Chenevix-Trench, G., 2006. What makes a good PhD student? Nature 441. https://doi.org/10.1038/nj7090-252b, 252-252.

Conti, A., Denas, O., Visentin, F., 2014. Knowledge specialization in Ph.D. student groups. IEEE Trans. Eng. Manag. 61, 52–67. https://doi.org/10.1109/TEM.2013.2283039.

Cornelissen, T., Dustmann, C., Schönberg, U., 2017. Peer effects in the workplace. Am. Econ. Rev. 107, 425–456. https://doi.org/10.1257/aer.20141300.

Cyranoski, D., Gilbert, N., Ledford, H., Nayar, A., Yahia, M., 2011. Education: the PhD factory. Nature 472, 276–279. https://doi.org/10.1038/472276a.

Dasgupta, N., Scircle, M.M., Hunsinger, M., 2015. Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. Proc. Natl. Acad. Sci., 201422822 https://doi.org/10.1073/pnas.1422822112.

Delamont, S., Atkinson, P., 2001. Doctoring uncertainty: mastering craft knowledge. Soc. Stud. Sci. 31, 87–107. https://doi.org/10.1177/030631201031001005.

Diamond, A.M., 1984. An economic model of the life-cycle research productivity of scientists. Scientometrics 6, 189–196. https://doi.org/10.1007/BF02016762.

Etzkowitz, H., 2003. Research groups as 'quasi-firms': the invention of the entrepreneurial university. Res. Policy 32, 109–121. https://doi.org/10.1016/S0048-7333(02)00009-4.

Falk, A., Ichino, A., 2006. Clean evidence on peer effects. J. Labor Econ. 24, 39–57. https://doi.org/10.1086/497818.

Fernandes, J.D., Sarabipour, S., Smith, C.T., Niemi, N.M., Jadavji, N.M., Kozik, A.J., Holehouse, A.S., Pejaver, V., Symmons, O., Bisson Filho, A.W., Haage, A., 2020.

A survey-based analysis of the academic job market. eLife 9, e54097. https://doi.org/10.7554/eLife.54097.
Freeman, R., Weinstein, E., Marincola, E., Rosenbaum, J., Solomon, F., 2001. Competition and careers in biosciences. Science 294, 2293–2294. https://doi.org/10.1126/science.1067477.
Gaulé, P., Piacentini, M., 2013. Chinese graduate students and U.S. scientific productivity. Rev. Econ. Stat. 95, 698–701. https://doi.org/10.1162/REST_a_00283.
Gaulé, P., Piacentini, M., 2018. An advisor like me? Advisor gender and post-graduate careers in science. Res. Policy 47, 805–813. https://doi.org/10.1016/j.respol.2018.02.011.
Golde, F.C.M., 2005. In: The Role of the Department And Discipline in Doctoral Student Attrition: Lessons From Four Departments, p. 33.
Heffernan, T., 2021. Academic networks and career trajectory: 'There's no career in academia without networks'. High. Educ. Res. Dev. 40, 981–994. https://doi.org/10.1080/07294360.2020.1799948.
Horta, H., Santos, J.M., 2016. The impact of publishing during PhD studies on career research publication, visibility, and collaborations. Res. High. Educ. 57, 28–50. https://doi.org/10.1007/s11162-015-9380-0.
Horta, H., Cattaneo, M., Meoli, M., 2018. PhD funding as a determinant of PhD and career research performance. Stud. High. Educ. 43, 542–570. https://doi.org/10.1080/03075079.2016.1185406.
Hunt, J., 2010. Why Do Women Leave Science And Engineering? (Working Paper No. 15853), Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w15853.
Katz, J.S., Martin, B.R., 1997. What is research collaboration? Res. Policy 26, 1–18.
Lee, S., Bozeman, B., 2005. The impact of research collaboration on scientific productivity. Soc. Stud. Sci. 35, 673–702. https://doi.org/10.1177/0306312705052359.
Lee, A., Dennis, C., Campbell, P., 2007. Nature's guide for mentors. Nature 447, 791–797. https://doi.org/10.1038/447791a.
Lempriere, D.M., 2020. What makes a good PhD supervisor?. In: Top Tips for Managing the Student-Supervisor Relationship, 13.
Levin, S.G., Stephan, P.E., 1991. Research productivity over the life cycle: evidence for academic scientists. Am. Econ. Rev. 81, 114–132.
Libaers, D.P., 2012. Time allocation decisions of academic scientists and their impact on technology commercialization. IEEE Trans. Eng. Manag. 59, 705–716. https://doi.org/10.1109/TEM.2012.2184119.
Liénard, J.F., Achakulvisut, T., Acuna, D.E., David, S.V., 2018. Intellectual synthesis in mentorship determines success in academic careers. Nat. Commun. 9, 4840. https://doi.org/10.1038/s41467-018-07034-y.
Long, J.S., McGinnis, R., 1985. The effects of the mentor on the academic career. Scientometrics 7, 255–280. https://doi.org/10.1007/BF02017149.
Lovitts, B.E., 2001. Leaving the Ivory Tower: The Causes And Consequences of Departure From Doctoral Study. Rowman & Littlefield Publishers.
Luckhaupt, S.E., Chin, M.H., Mangione, C.M., Phillips, R.S., Bell, D., Leonard, A.C., Tsevat, J., 2005. Mentorship in academic general internal medicine. J. Gen. Intern. Med. 20, 1014–1018. https://doi.org/10.1111/j.1525-1497.2005.215.x.
Ma, Y., Mukherjee, S., Uzzi, B., 2020. Mentorship and protégé success in STEM fields. Proc. Natl. Acad. Sci. U. S. A. 117, 14077–14083. https://doi.org/10.1073/pnas.1915516117.
Mangematin, V., 2000. PhD job market: professional trajectories and incentives during the PhD. Res. Policy 29, 741–756. https://doi.org/10.1016/S0048-7333(99)00047-5.
Mangematin, V., Robin, S., 2003. The two faces of PhD students: management of early careers of French PhDs in life sciences. Sci. Pub. Pol. 30, 405–414. https://doi.org/10.3152/147154303781780209.
Mas, A., Moretti, E., 2009. Peers at work. Am. Econ. Rev. 99, 112–145.
Merton, R.K., 1968. The Matthew effect in science. Science 159, 56–63.
Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space arXiv:1301.3781 [cs].
Moss-Racusin, C.A., Dovidio, J.F., Brescoll, V.L., Graham, M.J., Handelsman, J., 2012. Science faculty's subtle gender biases favor male students. Proc. Natl. Acad. Sci. 109, 16474–16479. https://doi.org/10.1073/pnas.1211286109.
Notman, N., Woolston, C., 2020. Fifteen to one: how many applications it can take to land a single academic job offer. Nature 584. https://doi.org/10.1038/d41586-020-02224-5, 315-315.
OECD, 2012. Closing the Gender Gap: Act Now. OECD Publishing, Paris.
OECD, 2013. Careers of Doctorate Holders: Analysis of Labour Market And Mobility Indicators (OECD Science, Technology And Industry Working Papers No. 2013/04). OECD Science, Technology and Industry. https://doi.org/10.1787/5k43nxgs289w-en. Working Papers.
OECD, 2019. Education at a Glance 2019: OECD Indicators [WWW Document]. URL. https://www.oecd-ilibrary.org/education/education-at-a-glance-2019_f8d7880d-en (accessed 2.4.21).
Overington, M.A., 1977. The scientific community as audience: toward a rhetorical analysis of science. Philos. Rhetor. 10, 143–164.
Paglis, L.L., Green, S.G., Bauer, T.N., 2006. Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. Res. High. Educ. 47, 451–476. https://doi.org/10.1007/s11162-005-9003-2.
Pearson, M., Brew, A., 2002. Research training and supervision development. Stud. High. Educ. 27, 135–150. https://doi.org/10.1080/03075070220119986c.
Pezzoni, M., Mairesse, J., Stephan, P., Lane, J., 2016. Gender and the publication output of graduate students: a case study. PLoS ONE 11, e0145146. https://doi.org/10.1371/journal.pone.0145146.
Platow, M.J., 2012. PhD experience and subsequent outcomes: a look at self-perceptions of acquired graduate attributes and supervisor support. Stud. High. Educ. 37, 103–118. https://doi.org/10.1080/03075079.2010.501104.
Pommier, S., Talby, M., Auffray-Seguette, M., Dalaut, M., Eijsberg, H., Elshawish, P., Muller, H., 2022. Le doctorat en France: Regards croisés sur la formation doctorale (Research Report). Réseau National de Collèges Doctoraux.
Powell, K., 2016. Does it take too long to publish research? Nat. News 530, 148. https://doi.org/10.1038/530148a.
Romano, J.P., Wolf, M., 2005a. Exact and approximate stepdown methods for multiple hypothesis testing. J. Am. Stat. Assoc. 100, 94–108. https://doi.org/10.1198/016214504000000539.
Romano, J.P., Wolf, M., 2005b. Stepwise multiple testing as formalized data snooping. Econometrica 73, 1237–1282.
Romano, J.P., Wolf, M., 2016. Efficient computation of adjusted p -values for resampling-based stepdown multiple testing. Stat. Probab. Lett. 113, 38–40. https://doi.org/10.1016/j.spl.2016.02.012.
Rong, X., 2014. word2vec Parameter Learning Explained arXiv:1411.2738 [cs].
Rossello, G., Cowan, R., Mairesse, J., 2020. Ph.D. Research Output in STEM: The Role of Gender And Race in Supervision, 57.
Sabatier, M., Carrere, M., Mangematin, V., 2006. Profiles of academic activities and careers: does gender matter? An analysis based on French life scientist CVs. J. Technol. Transfer. 31, 311–324. https://doi.org/10.1007/s10961-006-7203-3.
Sauermann, H., Haeussler, C., 2017. Authorship and contribution disclosures. Sci. Adv. 3, e1700404 https://doi.org/10.1126/sciadv.1700404.
Sauermann, H., Roach, M., 2012. Science PhD career preferences: levels, changes, and advisor encouragement. PLOS ONE 7, e36307. https://doi.org/10.1371/journal.pone.0036307.
Shibayama, S., 2019. Sustainable development of science and scientists: academic training in life science labs. Res. Policy 48, 676–692. https://doi.org/10.1016/j.respol.2018.10.030.
Shibayama, S., Kobayashi, Y., 2017. Impact of Ph.D. training: a comprehensive analysis based on a Japanese national doctoral survey. Scientometrics 113, 387–415. https://doi.org/10.1007/s11192-017-2479-7.
Shibayama, S., Baba, Y., Walsh, J.P., 2015. Organizational design of university laboratories: task allocation and lab performance in Japanese bioscience laboratories. Res. Policy 44, 610–622. https://doi.org/10.1016/j.respol.2014.12.003.
Sinclair, J., Barnacle, R., Cuthbert, D., 2014. How the doctorate contributes to the formation of active researchers: what the research tells us. Stud. High. Educ. 39, 1972–1986. https://doi.org/10.1080/03075079.2013.806460.
Stephan, P., 2006. Wrapping it up in a person: the mobility patterns of new PhDs. Innov. Policy Econ. 7, 71–98.
Stephan, P.E., 2012. How Economics Shapes Science. Harvard University Press, Cambridge, Mass.
Stephan, P.E., Levin, S.G., 1992. Striking the mother lode in science. In: The Importance of Age, Place, And Time. Oxford.
Stephan, P.E., Levin, S.G., 1997. The critical importance of careers in collaborative scientific research. Rev. Econ. Ind. 79, 45–61. https://doi.org/10.3406/rei.1997.1652.
Stephan, P., Black, G.C., Chang, T., 2007. The small size of the small scale market: the early-stage labor market for highly skilled nanotechnology workers. Res. Policy 36, 887–892. https://doi.org/10.1016/j.respol.2007.02.006.
Tartari, V., Perkmann, M., Salter, A., 2014. In good company: the influence of peers on industry engagement by academic scientists. Res. Policy 43, 1189–1203. https://doi.org/10.1016/j.respol.2014.02.003.
Taylor, A., Greve, H.R., 2006. Superman or the Fantastic Four? Knowledge combination and experience in innovative teams. Acad. Manag. J. 49, 723–740.
Tenenbaum, H.R., Crosby, F.J., Gliner, M.D., 2001. Mentoring relationships in graduate school. J. Vocat. Behav. 59, 326–341. https://doi.org/10.1006/jvbe.2001.1804.
Tompkins, K.A., Brecht, K., Tucker, B., Neander, L.L., Swift, J.K., 2016. Who matters most? The contribution of faculty, student-peers, and outside support in predicting graduate student satisfaction. Train. Educ. Prof. Psychol. 10, 102–108. https://doi.org/10.1037/tep0000115.
Vale, R.D., 2015. Accelerating scientific publication in biology. PNAS 112, 13439–13446. https://doi.org/10.1073/pnas.1511912112.
van Dijk, D., Manor, O., Carey, L.B., 2014. Publication metrics and success on the academic job market. Curr. Biol. 24, R516–R517. https://doi.org/10.1016/j.cub.2014.04.039.
Waldinger, F., 2010. Quality matters: the expulsion of professors and the consequences for PhD student outcomes in Nazi Germany. J. Polit. Econ. 118, 787–831. https://doi.org/10.1086/655976.
Waldinger, F., 2012. Peer effects in science: evidence from the dismissal of scientists in Nazi Germany. Rev. Econ. Stud. 79, 838–861. https://doi.org/10.1093/restud/rdr029.
Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. Science 316, 1036–1039. https://doi.org/10.1126/science.1136099.