Maastricht University

UNITED NATIONS UNIVERSITY

UNU-MERIT

Working Paper Series

#2021-011

What makes a productive Ph.D. student?

Alberto Corsini, Michele Pezzoni and Fabiana Visentin

Published 8 March 2021

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: **info@merit.unu.edu** | website: **http://www.merit.unu.edu**

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology
UNU-MERIT

What makes a productive Ph.D. student?

Alberto Corsini
Université Côte d'Azur, CNRS, GREDEG, France

Michele Pezzoni
Université Côte d'Azur, CNRS, GREDEG, France
Observatoire des Sciences et Techniques, HCERES, France
OFCE, SciencePo, France
ICRIOS, Bocconi University, Italy

Fabiana Visentin
Maastricht University, UNU-MERIT, The Netherlands

Abstract:

This paper investigates the impact of the social environment to which a Ph.D. student is exposed on her scientific productivity during the training period. Vertical and horizontal relationships depict the social environment. Vertical relationships are those supervisor-student, while horizontal relationships are those student-peers. We characterize these relationships by assessing how the supervisor's and peers' biographic and academic characteristics relate to the student's productivity as measured by the publication quantity, quality, and scientific network size. Unique to our study, we cover the entire student population of a European country for all the STEM fields. Specifically, we analyse the productivity of 77,143 students who graduated in France between 2000 and 2014. We find that having a female supervisor is associated with a higher student's productivity as well as being supervised by a mid-career scientist and having a supervisor with a high academic reputation. The supervisor's fundraising ability benefits only one specific dimension of the student's productivity, i.e., the student's work quality. Interestingly, the supervisor's mentorship experience negatively associates with student's productivity. Having many peers negatively associates with the student's productivity, especially if peers are senior students. Having female peers positively correlates with the student's productivity, while peers' academic status shows mixed effects according to the productivity dimension considered. We find results heterogeneity when breaking down our sample by field of research.

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

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). We thank Cornelia Lawson, Francesco Lissoni, Sofia Patsali, and Adrián Arias Diaz-Faes for their thoughtful comments. This work has been supported by the French government, 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.

"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."
(Testimonial by a second-year Ph.D. in Human Medicine)[1]

"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 is 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."
(Testimonial by a graduate student in Electrical engineering)[2]

1. Introduction

Nowadays, science increasingly relies on Ph.D. students' work. Ph.D. students play a fundamental role in advancing the scientific knowledge frontier with their publication activity (Larivière, 2012). Moreover, they are responsible for the knowledge flow among universities and between university and industry (Baruffaldi et al., 2020; Stephan, 2006). Recognizing the importance of highly skilled human capital, countries invest a relevant proportion of their total GDP in the higher education systems (OECD, 2019). Despite the importance of understanding the determinants of effective training programmes, a few studies have considered the early stages of scientists' formation (Shibayama, 2019). A large part of the literature has focused on experienced scientists (Carayol and Matt, 2006; Fox, 1983; Lissoni et al., 2011; Stephan, 1996).

The extant works on Ph.D. students' productivity have analysed only one productivity determinant at a time, focusing on the supervisor's gender, student affiliation quality, and scholarship funding (Conti et al., 2014; Gaule and Piacentini, 2018; Horta et al., 2018; Pezzoni et al., 2016; Waldinger, 2010). A first gap in the literature is that none of the extant studies has considered as productivity determinants the entire set of characteristics of the social environment in which the student is trained. Having a comprehensive overview of the impact of social environment characteristics is a fundamental subject of study since each of these characteristics might generate productivity differences during the Ph.D. period affecting the rest of the scientist's career (Allison et al., 1982; Azoulay and Lynn, 2020; Merton, 1968). A further characteristic of the extant works on Ph.D. students' productivity is that they focused on specific disciplines and relatively small samples of students affiliated to one or a handful of highly reputed universities. A second gap in the literature

[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

is that none of the extant studies have conducted empirical analysis on the entire population of Ph.D. students of a country, including students enrolled both in top-tier universities and low-rank universities as well as students belonging to different research fields.

We fill these two gaps by analysing the impact of the social environment to which a Ph.D. student is exposed during her training period on her scientific productivity. Unique to this study, we analyse the entire population of STEM graduates from one European country, France, over fifteen years. As social environment, we consider the set of vertical and horizontal relationships established by the Ph.D. student during the training. Vertical relationships are between the student and the supervisor, while horizontal relationships are between the student and her peers. The supervisor plays the mentor's role and transfers knowledge and skills to her students (Shibayama, 2019; Stephan and Levin, 2002). Although students refer to their supervisors, the learning process is largely affected by group dynamics. Students spend most of their time in labs or classes and frequently interact with their peers. Both the characteristics of supervisors and peers relate to effective learning. This paper considers supervisor's and peers' biographic and academic profile as social environment relevant characteristics.

Concerning the biographic characteristics, we find that both supervisor's and peers' gender is weakly associated with the student's publication quantity, quality, and network size. While the supervisor's seniority shows an inverted U-shaped relationship with all the student's productivity dimensions, peers' average seniority is associated with a decline in the student's productivity. Regarding the academic characteristics, we find that a one-standard-deviation increase in supervisors' publications is associated with 0.39 additional students' publications, while an increase of the same extent of the peers' publications is associated with 0.23 additional students' publications. Looking at students' publication quality and co-authorship network size, we find that they are positively associated with the supervisor's productivity while finding mixed evidence on their association with peers' productivity. Interestingly, both national and European grants awarded to the supervisor are associated with an increased student work quality as measured by the citations received. Student's work receives 0.54 additional yearly citations if supervised by a researcher who benefitted from a national grant and 0.33 citations if supervised by a researcher who benefitted from a European grant. When we break down our analysis by field, i.e., Mathematics, Engineering, Physics, and Medicine-biology-chemistry, we find result heterogeneity across fields. All our econometric estimates control for the student's characteristics and for the characteristics of the department where the student is enrolled.

2. Vertical and horizontal relationships during the training period: Impact on Ph.D. students' scientific outcomes

As in any other working context, when students start their Ph.D. training, they become part of a social environment characterized by vertical and horizontal relationships. Vertical relationships are between the student and her supervisor, while horizontal relationships are between the student and her peers. This section provides a theoretical framework to illustrate how vertical and horizontal relationships characterizing the social environment affect students' productivity.

Vertical relationships: Supervisor's characteristics and student's productivity

Vertical relationship dynamics affect the Ph.D. experience (Chenevix-Trench, 2006; Lempriere, 2020). These dynamics are regulated by an implicit contract between the student and her supervisor (Mangematin, 2000; Stephan and Levin, 2002). In this contract, the student contributes to the supervisor's scientific productivity with her work, time, and effort, while the supervisor helps the student to complete the training programme, transferring scientific competencies, and offering access to her scientific networks and resources (Long and McGinnis, 1985; Platow, 2012).

The biographic and academic characteristics of the supervisor affect the successful outcome of the implicit contract. Looking at the biographic characteristics, previous literature has investigated how the supervisor's gender affects student's productivity during the Ph.D. training period. In chemistry, Gaule 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. In the context of a US leading interdisciplinary university, Pezzoni et al. (2016) find that having a female supervisor increases Ph.D. students' productivity. Interpreting these empirical results involves sociological aspects at the root of the different mentoring approaches adopted by female and male supervisors. Surveying 185 students at the University of California, Tenenbaum, Crosby, & Gliner, (2001) find that male supervisors are less likely than their female counterparts to provide psychological help to the students decreasing their level of satisfaction with the Ph.D. training experience. However, both female and male supervisors offer equal "instrumental help," providing students the same technical knowledge to enhance their publication productivity.

Another characteristic that affects students' productivity is the supervisors' seniority. Students might benefit from having young supervisors. As suggested by the labour literature, a rational individual decreases the time devoted to working with seniority (Diamond, 1984; Levin and Stephan, 1991). Moreover, scientists have more autonomy than in other jobs in choosing the time allocation to different activities such as fundraising, research, teaching, consulting, and administrative activities (Libaers, 2012; Sabatier et al., 2006). By combining these two characteristics of the academic job, we expect time allocation to different activities to evolve with seniority. Young supervisors aiming to boost their careers tend to devote their time to fundraising, research, and mentoring activities. In contrast, senior supervisors are likely to dedicate more time to activities more remunerative in the short-term such as consulting and administrative activities. Less time spent in research and supervision by a senior supervisor might impact negatively on the support provided to her Ph.D. students, and ultimately on the students' productivity.

Although the supervisor's seniority and mentorship experience positively correlate, they are two different concepts that are expected to have opposite effects. Not only students but also supervisors learn from mentoring. Supervising experience develops different abilities, such as advising, tutoring, encouraging, and providing a role model to students (Broström, 2019). The supervisor's mentoring skills might evolve with experience and lead to better training of the student when the supervisor has a long history of mentored students. This better training is expected to be associated with the higher productivity of the Ph.D. student.

Considering the supervisor's academic characteristics, publication and citation productivity reflect her 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). 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 the successful behaviour of productive supervisors increases the student's probability of showing a high productivity level during the Ph.D. period.

For an early-stage scientist, building a strong publication record is as important as establishing a network of co-authors to leverage the future career development. Indeed, teamwork has become a requirement to have a productive scientific career (Börner et al., 2010; Wuchty et al., 2007). One of the most important contributions of the supervisor to the student's productivity is to help the student creating her scientific collaboration network (Long and McGinnis, 1985; Tenenbaum et al., 2001). Students supervised by scientists in contact with many colleagues are more likely to spend visiting periods in other labs acquiring new competencies, to be introduced to leading scientists in the discipline, and to be exposed to different research approaches (Mangematin and Robin, 2003; Stephan, 2006). These networking opportunities are expected to positively impact the student's productivity (Lee and Bozeman, 2005).

Besides publication and networking influence, supervisors are fundamental in providing resources that contribute to the successful completion of their students' Ph.D. programme. The funds made available by the supervisor's fundraising ability directly affect the student's productivity.

Recently scholars have focused on the impact of different types of scholarships on students' productivity (Horta et al., 2018). Although French Ph.D. scholarships can be funded by universities, laboratories, or the State, every student's hiring contract is relatively standard (Mangematin, 2000). Thus, the difference in the availability of financial and material resources to students does not depend on the type of scholarship but on the supervisor's fundraising ability. Indeed, the student's conference participation, the possibility of visiting periods, and the lab equipment are primarily funded by the budget of the competitive grants awarded to the supervisor. Therefore, the supervisor's abundance of research funding might significantly affect the Ph.D. student's productivity and working conditions.

Horizontal relationships: Peers' characteristics and student's productivity

We define horizontal relationships as the student's relationships with peers. We define 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).

Ph.D. students, as any other worker, interact with peers during their professional activity. These interactions might affect students' productivity in two ways. On the one hand, students feel the "peer pressure" of maintaining a level of productivity similar to that of their peers striving for scientific recognition by their supervisor and the scientific community (Stephan and Levin, 1992). On the other hand, students might learn by observing and interacting with their peers (Ayoubi et al., 2017; Cornelissen et al., 2017). The learning process might regard practical issues such as the best strategies to obtain the supervisor's attention, financial resources, and lab equipment use. Peers' interaction might also stimulate scientific discussions leading to knowledge acquisition from peers and generating novel research ideas (Ayoubi et al., 2017). Both peer pressure and learning from peers are mechanisms expected to increase the student's productivity.

Labour literature, both using observational data and experimental data, is convergent in showing that having peer co-workers in the work environment positively affects productivity (Falk and Ichino, 2006). Although the expected beneficial effect of having peers, working in a research environment characterized by large groups might generate coordination costs and competition dynamics that negatively affect the students' outcomes (Broström, 2019). Moreover, the supervisor's time allocated to each student might reduce when the number of students increases. Therefore, we expect that the beneficial effect of having peers shrinks when the peers' number increases.

Not only the mere presence of peers is expected to affect the focal student's productivity, but also peers' characteristics. Similarly to the supervisor, we consider peers' biographic and academic characteristics. Concerning the biographic characteristics, Dasgupta et al. (2015) find that group dynamics are not gender-neutral. Specifically, they find that female students' participation in group discussions and self-confidence are higher in female-majority groups, both for junior and senior group members. Thus, in the Ph.D. training context, we expect the gender composition of the peer groups to influence the Ph.D. student's behaviour and, ultimately, her productivity. Similarly, we expect the peers' seniority to affect the student's productivity. On the one hand, having more senior peers with greater knowledge stocks enhances knowledge transfer toward the focal student (Ayoubi et al., 2017). This knowledge transfer toward the student might increase her productivity. On the other hand, more senior peers might be in a phase of their Ph.D. when ideas are already settled, leading to less creative interactions with the focal student.

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, i.e., 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 the not convergent results, in the Ph.D. students' context, we expect that highly productive peers will boost the student's productivity, both through the mechanisms of "peer pressure" and to the enhanced probability of acquiring knowledge from productive peers.

Peers might play a role also 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.

3. French Ph.D. students in the STEM field

Our empirical setting is represented by STEM French Ph.D. students. France is one of the European countries excelling in the STEM field. Looking at the absolute number of Nobel Prize winners, 39 French scientists obtained the highest recognition in the fields of 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 women who obtained a Nobel Prize and the only women 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. All French universities and *Grandes écoles* can hold their own Ph.D. programmes. Ph.D. students in natural and technological sciences work full time in research labs with their colleagues, while in the other disciplines their work does not require a daily basis presence in labs. Ph.D. students are asked to attend core classes in theory and methodology and additional skill classes such as writing scientific papers. A considerable amount student's Ph.D. time is dedicated 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 career progress. The final thesis importance is evident from the fact that French people often interchange the expression "being enrolled in a Ph.D. programme" with "*faire une these*" (the English equivalent of "writing a thesis"). To access the doctoral programme, candidates need to be paired with a thesis supervisor who accepts to guide them. The most common way of completing a Ph.D. is writing a thesis under the guidance of a single supervisor; however, co-supervised doctorates are possible.

4. Data

To construct our study sample, we gather data from multiple sources. The first is the nation-wide French repository of Electronic Doctoral Theses (EDT). 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 on the author, the university of graduation, the defence date, the supervisor name, the co-supervisor name (if any), and the field of study. As fields, we distinguished theses in Mathematics, Engineering, Physics, and Medicine-biology-chemistry. The records do not report the year of entry of the student into the Ph.D. programme; thus, we approximate it assuming that each student started three years before her thesis defence year. According to the national statistics for STEM fields, the most frequent duration of the Ph.D. training in France is four years, including the defence year[3]. Hence, we define the *Ph.D. training period* as the period ranging from three years before the defence year to the defence year *t*, i.e., from *t-3* to *t*. According to this definition, we set the student's *entry year* into the Ph.D. programme as the first year of the training period, i.e., *t-3*.

Our information on the students' and supervisors' gender results from a multiple-iteration matching strategy (Gaule and Piacentini, 2018; OECD, 2012). First, we match the given names with the official French gender-name dataset[4]. Then, for the non-matched names, we repeated the matching

[3] We double checked this statistic by querying the universities' administration.
[4] Website: https://www.data.gouv.fr/fr/datasets/liste-de-prenoms/

exercise with the *U.S. Census Bureau* gender-name dataset and with WIPO gender-name dataset[5], respectively.

We retrieve students' and supervisors' publication records from Elsevier's SCOPUS database.

We gather information on funding at the national as well as 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 programmes 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.

To reconstruct the quality of the Ph.D. students' graduation department, we rely on the QS university ranking[6]. The QS university ranking provides detailed information on the university academic reputation at the department level and allowed us to flag the top departments for each field. For instance, *Université de Paris* is in the top-20 percent of universities in Mathematics in France, but not in Engineering. We integrate the information on QS ranking with bibliometric information concerning the university affiliates and constructing an appropriate bibliometric indicator at the department level. As an additional proxy for the department quality, we identify the French universities that in 2011 benefitted from the *Initiative D'Excellence* (IDEX) "block" funding provided by the French Government to a selected group of French higher education institutions. The IDEX funding programme was launched in 2011 by the French Government within a national fiscal stimulus and awarded to eight universities[7] that strive to become competitors of worldwide top-ranked universities.

To create our study sample, we joined all the information described above using the student's full name as the merging criterion. Then, we refined our study sample excluding homonym students[8], students with more than 20 publications and students with more than 100 citations received per paper during the Ph.D. period, being the productivity of these latter too high to be credible. Overall, the excluded students represent around 7% of our initial sample. After this cleaning exercise, we obtain a study sample of 77,143 Ph.D. students who graduated between 2000 and 2014 from French universities.

When classified by field, 15% of the students are in Mathematics, 18% in Physics, 21% in Engineering, 45% in Medicine, Biology, and Chemistry. The students publish on average 2.37 peer-

---

[5] Website: https://www.wipo.int/publications/en/details.jsp?id=4125
[6] Website: https://www.topuniversities.com
[7] 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.
[8] Having two of more students with the same full name in our original list of Ph.D. thesis authors would make 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.

reviewed articles during their training period. 68% percent of students publish at least one article during the Ph.D. period. The average students' collaboration network includes 8.93 distinct co-authors over the training period.

The average supervisor has a stock of 13.59 peer-reviewed articles and a seniority of 11.49 years of career when her student has enrolled in the Ph.D. programme. At the time of the student's enrolment, the average supervisor counts 3.08 successfully supervised Ph.D. students over her career. While the percentage of students doing a Ph.D. in STEM does not dramatically differ by gender, 39% are women and 61% are men, only 21% of the supervisors are women. When considering the funding, only 6% of the students have a supervisor who is the principal investigator of an ANR national grant during the Ph.D. training period. Only 2% of the students have a supervisor who is the principal investigator of an EU grant.

We identify the focal Ph.D. student's peers as those students having an overlapping training period with the focal student and having the same supervisor of the focal student. 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.

5. Econometric methodology

To estimate the impact of the vertical and horizontal relationships characterizing the Ph.D. student's social environment on her productivity, we estimate the coefficients of the model presented in Equation 1 using *Ordinary Least Squares* (OLS). The level of analysis, as represented by the subscript *i*, is the student.

$$Student's\ productivity_i = \\ \beta_0 + \beta_1\ Supervisor's\ characteristics_i + \beta_2\ Peers'characteristics_i + \beta_3\ Controls_i + \varepsilon_i$$

Equation 1

The left-hand side variable in Equation 1 takes, in turn, the value of the student's publication quantity, quality, and 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 defence 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 defence year.

The vectors *Supervisor's characteristics* and *Peers' characteristics* define the characteristics of the vertical and horizontal relationships in the Ph.D. student's social environment. *Controls* is a vector including the student's characteristics and the characteristics of the university where the student is enrolled. Finally, ε is the idiosyncratic error term.

A potential concern in the estimation of the variable coefficients in our econometric model relates to a possible endogeneity issue. The lack of proxies for the student's intrinsic ability might result in biased estimates of the coefficients if the unobserved ability correlates with the 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. Previous studies (Mangematin, 2000) 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. In other words, asymmetry of information in students' selection makes unluckily to observe a correlation between students' intrinsic ability and supervisors' quality. Moreover, 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 outcomes and training, are uncorrelated with the student's productivity. Nonetheless, in Section 6.2, we implement a robustness check to respond to the potential endogeneity concern. We replicate the estimations of Equation 1 adding a proxy that controls for the ability of the student during her high school period. We flag students with exceptional quality by constructing a dummy variable equal to one if the student participated in a selective contest during her last years of 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 before starting her graduate studies 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.

Vertical relationships: Supervisor's characteristics

To characterize the vertical relationships between the student and the supervisor, we consider the supervisor's biographic and academic characteristics.

Concerning the biographic characteristics, we include a dummy variable *Female supervisor* which 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. programme. 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 who have successfully defended their thesis coached by the supervisor along her career[9].

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. programme, i.e., from *t-4* to *t-8*, where *t* is the student's defence year. We decided to measure the supervisor's publication quantity and quality during the five years preceding the student enrolment (and not during the student training period) since it is a common practice that the student and her supervisor co-sign publications during the student's training period. In the case of co-signed articles, it is impossible to disentangle the supervisor's productivity from the 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. programme. Then, we calculate, for the same period, 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 had in the five years preceding the student's entry into the Ph.D. programme. Finally, to proxy for the supervisor fundraising ability, we calculate a dummy *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 *EU grant* that equals one if the supervisor is the principal investigator of an EU grant during the student's training period.

Horizontal relationships: Peers' characteristics

Ph.D. students spend their training periods in a social environment, either with or without peers. To characterize the presence of peers in the social environment, 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. We calculate the variable *N. peers* as the yearly number of students with whom the focal student shares the training experience. To account for the fact that peers might have only partially overlapping training periods with that of the focal student, we first calculate the yearly number of peers in each of the four years of the focal

[9] We retrieve data on supervisors' coaching career from 1980.

student's training period; then, we obtain the variable *N. peers* averaging the four values. For instance, if the focal student spends the first three years of her training period without peers and then her supervisor recruits another student in the last year of the focal student's training period, the variable *N. peers* for the focal student takes the value of 0.25 (0.25=(0+0+0+1)/4).

To characterize the student's relationships with 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. programme (*Average peers' seniority*). Also in this case, peers might have only partially overlapping training periods 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. In case 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 values. For instance, if the focal student has only one peer during her training period, and that peer defends the thesis during the second year of the focal student's training period, the peer's seniority equals 3 and 4 during the two overlapping years. Therefore, 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[10] 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* and the variable *Peers' co-authors*.

Other controls

To mitigate the potential bias of our estimated coefficients, we control for the characteristics of the department in which the student is enrolled and for the student's characteristics. We define a department as the pair university-field. For instance, *Université de Paris* counts four departments:

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

Université de Paris Mathematics, *Université de Paris* Engineering, *Université de Paris* Physics, and *Université de Paris* Medicine-biology-chemistry.

To control for the department quality, we retrieve the university reputation ranking from the QS World University ranking[11]. We create a dummy *French Top-20* that equals one if the department is among the 20% 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 variable, we consider the department affiliates' average productivity during the five years preceding the student's entry into the Ph.D. programme. Specifically, we identify the department affiliates' publications during the five years preceding the student's enrolment. 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 and awarded with the IDEX national investment programme funding.

To control the department size, we calculate the variable *Department size* counting the number of scientists affiliated with the department for at least one year during the five years preceding the student's entry into the Ph.D. programme[12]. We rescale the number of affiliates dividing by 100, meaning that each unit increase of the variable *Department size* corresponds to 100 additional department affiliates.

Along with the department size, the size of the Ph.D. programme might play a role. Larger Ph.D. programmes might be better organized and provide the student with a better and productive training experience. We calculate the number of Ph.D. students enrolled in the focal student's Ph.D. programme for each of the four years of her training period. Then, we calculate the variable *N. of Ph.D. students in the programme* 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[13]. 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*

[11] 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.

[12] We retrieve the scientists' affiliation from their publications.

[13] 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.

controlling for the non-observable heterogeneity across the thesis research fields. Finally, we add a set of dummy variables for the students' *Entry year* to account for the Ph.D. cohort effect.

Table 1 lists all the variables included in our analysis with a short description for each of them. Table 2 reports the descriptive statistics for the variables calculated on our sample of 77,143 Ph.D. students.

Table 1. List of variables used in the analysis.

	Variable description
Dependent variables	
<u>*Student's productivity*</u>	
Publications	Ph.D. student's number of papers published between t-3 and t+1*
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	
<u>*Supervisor characteristics*</u>	
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 an EU grant between t-3 and t
<u>*Peer characteristics*</u>	
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. programme 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
<u>*Other controls*</u>	
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 programme	Average number of Ph.D. students per year enrolled in the focal student's Ph.D. programme 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. programme, i.e., t-3

NOTE: *t is the Ph.D. thesis defence year; t-3 is the entry year of the student into the Ph.D. programme; 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. programme.

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*
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 programme	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

NOTE: *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.

6. Results

Table 3 reports the OLS estimates of the model described in Equation 1.

Table 3. Regression results. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor characteristics			
Female supervisor	-0.0051	0.074**	0.31**
Supervisor's seniority	0.037***	0.0071	0.11***
Supervisor's seniority^2	-0.0019***	-0.00096**	-0.0067***
Mentorship experience	-0.018***	-0.0072***	-0.037***
Supervisor's publications	0.027***	0.0070***	-0.10***
Supervisor's average citations	0.031***	0.20***	0.21***
Supervisor's co-authors	0.0028***	0.0014***	0.091***
ANR grant	0.0048	0.54***	0.22
EU grant	-0.19***	0.33***	-1.28***
Peer characteristics			
With peers	0.13***	0.24***	0.25
N. peers	-0.12***	-0.042***	-0.39***
At least one female peer	-0.028	0.073**	0.21*
Average peers' seniority	-0.14***	-0.13***	-0.63***
Peers' publications	0.13***	-0.15***	-0.64***
Peers' average citations	0.0065***	0.056***	0.049***
Peers' co-authors	0.0029	0.0017	0.21***
Other controls			
French Top-20	-0.0082	0.068**	-0.36***
Citation-weighted publications per affiliate	0.012**	0.026***	0.14***
IDEX	-0.056	0.031	-0.032
Department size [100 affiliates]	0.00081	0.0014**	0.013***
N. of Ph.D. students in the programme	0.000092***	0.00023***	0.00038***
Female student	-0.64***	-0.19***	-1.84***
Co-supervision	-0.066***	-0.042	-0.66***
Engineering	0.18***	0.40***	0.99***
Physics	0.77***	0.57***	2.46***
Medicine-biology-chemistry	1.54***	1.39***	6.45***
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	1.23***	0.23***	3.83***
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.

Looking at the impact of the supervisor's biographic characteristics on the student's productivity, we find that having a *Female supervisor* is not associated with the number of papers published by the student. 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). Although statistically significant, these two variations are economically limited, corresponding to the 3.5%[14] of the sample

[14] 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).

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 all the three student outcomes considered. The maximum impact of seniority on the student's publication productivity, citations, and network size is when the supervisor has 9.74[15], 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 has smaller collaboration networks. Specifically, increasing by one standard deviation, the *Mentorship experience* is associated with 0.11 fewer papers, 0.045 fewer citations, and 0.23 fewer co-authors. Although statistically significant, these variations are limited compared to the means of the three dependent variables in our sample, corresponding to 4.64% of the student's average publication productivity, 2.13% of the average citations, and 2.58% of the average number of co-authors. This result can be interpreted as the supervisors' tendency to be more supportive to the student when they are at the first experiences as thesis directors.

Looking at the supervisor's academic characteristics, supervisor's productivity, i.e., *Supervisor's publications*, *average citations*, and *co-authors*, is associated with a higher student's productivity. Specifically, increasing the supervisor's publication by one standard deviation is associated with 0.39[16] additional student publications (16.3% of the sample average[17]) and 0.10 additional citations (4.75% of the sample average). Similarly 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). Overall, our results show a positive relationship between the supervisor's academic characteristics and the productivity of the Ph.D.

[15] 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).
[16] 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.
[17] 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).

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. This time is probably subtracted from mentoring the student. Although we observe some differences between national 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. of peers*, since, when the dummy variable *With peers* equals one, the value of the variable *N. of peers* is a positive integer number. Therefore, we find that the overall effect of having one peer only is associated with 0.20 (=0.24-0.042) additional citations (9.4% of the sample average) and we do not observe any statistical significance[18] of having one peer for the publication quantity and co-authorship network size. Although having one peer is associated with benefits to productivity quality, we find that 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. These three values correspond to the 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.

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. Although statistically significant, the increase in the student’s citations and co-authors is limited to 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-

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

authors (7.34% of the sample average). These results lead us to conclude that peers' gender has limited positive relationships with the student's productivity, while peers' seniority negatively associates 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.

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 programme 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. programme (*N. of Ph.D. students in the programme*) is positively associated with all the Ph.D. student's productivity dimensions. Larger Ph.D. programmes 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)[19]. Moreover, the presence of a co-supervisor is detrimental to the student's productivity outcomes.

[19] 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.

Looking at the set of dummies identifying the fields of study, we observe productivity heterogeneity across fields. The latter result is expected since different fields are characterized by different 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 coefficients of Equation 1 for students in Mathematics, Engineering, Physics, and Medicine-biology-chemistry.

6.1 Exploring heterogeneity across fields

A possible concern in exploring the determinants of Ph.D. students' outcomes is cross-field heterogeneity. Supervisor's and peers' characteristics might have a different impact on the students' productivity. In this section, we dig into the field heterogeneity by conducting separate analysis by field. Table 4 reports the statistics of Ph.D. students' productivity by field. 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 the field of 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 A.1 in the Appendix reports the descriptive statistics of the complete set of explanatory variables by field.

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

Dependent Variables	Engineering	Mathematics	Medicine-Biology-Chemistry	Physics
Observations	16,519	11,450	35,038	14,136
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

Table 5 reports the estimations of the coefficients of Equation 1 by field. Looking at the supervisors' biographic characteristics, differently from our main regressions presented in Table 3, the relationship between the supervisor's seniority and the student's productivity is not statistically significant in Engineering and Physics. Having a female supervisor relates positively to students' productivity in Engineering, while the effect is limited in the other fields. A female supervisor in Engineering benefits the Ph.D. student with 0.25 additional publications, 0.29 yearly citations, and 0.79 co-authors. The supervisor's mentorship experience shows the same association with all the student's outcomes across fields: the greater the number of students mentored in the past by the supervisor, the lower the student's productivity outcomes.

When looking at the supervisors' academic characteristics, having a strong publication profile has a homogeneous 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 supervisors' 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 the Ph.D. student's productivity is positive in Medicine-biology-chemistry, while it is negative in the other fields.

Results reported in Table 5 show that being mentored by a supervisor who benefited from an ANR grant is positively associated with the Ph.D. students' overall productivity in Engineering and Physics. When we consider European grants, instead of national grants, we find that EU grants are positively associated with students' citations in Physics and Medicine-biology-chemistry. This latter result might be explained by the high student visibility gain in these fields due to the collaboration with other European countries.

In all fields, the increase in the number of peers is associated with decreased student's productivity, with the sole 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 one female peer is associated with productivity benefits in all the disciplines, except in Physics, where having a female peer is negatively associated with the Ph.D. students' publication productivity (-0.17 publications).

Peers' academic characteristics show mixed effects on student's productivity outcomes. Interestingly, the peers' network size is particularly favourable 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.

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**
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
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
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**
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***
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***
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***
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***
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
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**
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***
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
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***
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
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
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***
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
Citation-weighted publications per affiliate	0.035	0.00029	0.11	0.042**	0.064**	0.15**	-0.035*	0.017	-0.084	0.0044	0.021*	0.064
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
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***
N. of Ph.D. students in the programme	0.00013***	0.000022	0.00066***	0.00013***	0.000082**	0.00030***	-0.000076***	0.00028***	-0.00022	0.00019***	0.00038***	0.0011***
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***
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
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***
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.

6.2 Robustness checks

As mentioned in Section 5, we cleaned our dataset by excluding homonym students, students with more than 20 publications, and students with more than 100 citations received per publication during the training period. In doing so, we expect to mitigate the wrong attribution of publications to Ph.D. students when using the student's name as the key variable. Nevertheless, some erroneous attributions could still be present in our analysis. In this section, we propose two robustness checks using two alternative methods to attribute publications to students. First, we attribute to the student only publications listing among the authors both the student's name and the supervisor's name. Second, using a text analysis algorithm to assess the thesis and publication content, we attribute to the student only publications listing among the authors the student's name and having similar content to her thesis manuscript.

In Appendix 2, Table A.2. reports the descriptive statistics of the student's productivity variables calculated considering only the publications co-authored by the student with the supervisor. On average, we find that a Ph.D. student publishes 1.76 papers co-authored with the supervisor, receives 1.97 yearly citations per paper, and has a network of 6.92 co-authors during the training period. Table A.3. shows the regression estimates of Equation 1 using the three newly calculated dependent variables. Although results are largely consistent with the ones reported in our main analysis in Table 3, some results differ. Specifically, having a supervisor who is the principal investigator of an ANR grant is positively associated with all three productivity outcomes and not only with the productivity quality as in Table 3. In particular, having a supervisor awarded an ANR grant is associated with 0.16 additional publications, 0.58 additional yearly citations, and 0.71 additional co-authors.

In Appendix 3, Table A.4 reports the descriptive statistics of the three dependent variables calculated considering only the publications authored by the student which are similar to the student's thesis manuscript. To measure the similarity between the publications authored by the student and a student's thesis, we rely on a text analysis algorithm that compares the abstracts of the publications with the abstract of the thesis (Mikolov et al., 2013). According to this attribution method, we find that, on average, a student publishes 1.38 papers, receives 1.41 yearly citations, and has 5.37 co-authors during the training period. Table A.5 reports the regression estimates of Equation 1. The regression results are largely consistent with the ones reported in Table 3. There are only two exceptions. The first exception regards the relationship between the supervisor's seniority and the Ph.D. student's productivity, which now turns into a U-shaped relationship with the student's productivity. Second, having a supervisor who is the principal investigator of an ANR grant positively correlates with all the student's productivity outcomes.

In a further regression exercise, in Appendix 4, we construct a variable proxying for the student's intrinsic ability. To do so, we collected data on 138 students who participated in three national and international Mathematical Olympiad-like contests during their high school studies. We define the dummy variable *Math Olympiad* as a variable that equals one if the students participated in at least one of the contests, zero otherwise. We find that when we include *Math Olympiad* in our regression exercises, the estimated coefficients of the variables of the supervisor's and peers' characteristics are in line with those reported in Table 3.

7. Conclusion

Ph.D. students nowadays are considered key players in the scientific knowledge production process. Their productivity during the training period is an essential contribution to the advancement of the scientific frontier (Halse and Mowbray, 2011; Larivière, 2012).

In this paper, we study how the social environment influences the Ph.D. students' productivity during their training period using a dataset that considers the entire population of 77,143 Ph.D. students who graduated from French universities in STEM disciplines between 2000 and 2014. For the first time, we consider the entire Ph.D. population of a European country in a large set of disciplines. As relevant dimensions of the social environment, we consider the vertical relationships between the student and the supervisor and the horizontal relationships between the student and the peers. To characterize these relationships, we do look at the biographic and academic characteristics of supervisors and peers. To measure the student's productivity during the training period, we consider three productivity dimensions: publication quantity, publication quality, and size of the scientific network. Publication quantity counts the number of articles published, publication quality is calculated as the average number of citations received per article, and the scientific network size equals the number of distinct co-authors.

We find that supervisors' biographic and academic characteristics influence students' productivity. Having a female supervisor is associated with an increase of 0.074 citations and 0.31 co-authors, respectively. We also find that working with too young or too senior supervisors is detrimental to the student's productivity for all the three productivity dimensions considered. This result shows that mid-career supervisors are associated with better student outcomes. When the supervisor is in the late-career stages, the student's lower productivity can be explained by a higher supervisor's commitment toward non-research activities, such as administrative and teaching activities. As expected, having a productive supervisor is associated with a higher student's productivity. A one-standard-deviation increase in supervisors' publications is associated with 0.39 additional student publications. Supervisor's citations and supervisor's number of co-authors positively correlate with all the student's productivity dimensions. The only exception regards the relationship between the supervisor's number of publications and the student's network size: increasing the supervisor's publications by one standard deviation is associated with 1.43 fewer student co-authors. Interestingly, the mentorship experience is detrimental to the Ph.D. student's productivity. Although of limited size, this result might suggest that supervisors at their first mentoring experiences devote more effort to support their students than experienced supervisors. The supervisor's availability of French research grants does not correlate with the student's publications, while European funds negatively correlate. Both national and European funds positively correlate with the students' citations received. The positive results on citations for both national and EU grants might be interpreted as increasing the research group's visibility due to the awarded grants. European funds are also detrimental to the network size. The negative correlations between the European funds and students' publication quantity and network size could be explained by an increase in the administrative burden required by these grants that forces the supervisor to allocate less time to the mentoring activity.

Sharing the training experience with large groups of peers penalizes productivity, showing that when the supervisor has too many students, the quality of the mentoring activity declines. Peers' biographical and academic characteristics matter. Having at least one female peer is positively associated with student's citations and network size, although the increase in the student's citations and co-authors is limited to 0.07 citations and 0.21 co-authors. Having freshman peers relates positively to the students' productivity, as well as having productive peers. Peers' publication productivity is positively associated with the student's publications but negatively with her citations and co-authorship network size. A one-standard-deviation increase in peers' publications is associated with 0.23 additional students' publications, 0.26 fewer citations, and 1.13 fewer co-authors. An increase in peers' average citations is positively associated with the student's productivity. An increase in peers' co-authors benefits only the focal student's network size.

We dig into the non-convergent findings of the previous literature on the determinants of students' productivity by exploring field heterogeneity. We interpret these not convergent findings as the result of field specificities. For instance, aligned with Waldinger (2010), we show a positive influence of the department's prestige on Ph.D. students' productivity in Mathematics. However, we show that this result does not hold for students in Engineering and Medicine-biology-chemistry.

Our results talk to both Ph.D. students and policymakers. On the one hand, Ph.D. students are facing nowadays a highly competitive job market after graduation. Especially those who want to pursue an academic career need to show a high-quality publication record and have a well-established scientific network. Our paper provides hints to the students who want to leverage the environmental factors to develop their productivity. On the other hand, our results provide the policymakers with a framework to understand the determinants of effective training programmes and find levers to design policies that maximize students' productivity. For instance, our results show that, all else equal, limiting the number of Ph.D. students mentored at the same time by a supervisor benefits the students' productivity. Moreover, favouring the Ph.D. supervision of mid-career scientists benefits the student's productivity.

References

Agarwal, R., Gaule, P., 2020. Invisible Geniuses: Could the Knowledge Frontier Advance Faster? American Economic Review: Insights 2, 409–424. https://doi.org/10.1257/aeri.20190457

Allison, P.D., Long, J.S., Krauze, T.K., 1982. Cumulative advantage and inequality in science. American Sociological Review 47, 615–625.

Ayoubi, C., Pezzoni, M., Visentin, F., 2017. At the origins of learning: Absorbing knowledge flows from within the team. Journal of Economic Behavior & Organization.

Azoulay, P., Lynn, F.B., 2020. Self-Citation, Cumulative Advantage, and Gender Inequality in Science. Sociological Science 7, 128–151. https://doi.org/10.15195/v7.a7

Azoulay, P., Zivin, J.S.G., Wang, J., 2010. Superstar extinction. Quarterly Journal of Economics 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.H., Marino, M., Visentin, F., 2020. Money to move: The effect on researchers of an international mobility grant. Research Policy 49, 104077.

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

Börner, K., Contractor, N., Falk-Krzesinski, H.J., Fiore, S.M., Hall, K.L., Keyton, J., Spring, B., Stokols, D., Trochim, W., Uzzi, B., 2010. A Multi-Level Systems Perspective for the Science of Team Science. Science Translational Medicine 2, 49cm24-49cm24. https://doi.org/10.1126/scitranslmed.3001399

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

Carayol, N., Matt, M., 2006. Individual and collective determinants of academic scientists' productivity. Information Economics and Policy 18, 55–72. https://doi.org/10.1016/j.infoecopol.2005.09.002

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

Conti, A., Denas, O., Visentin, F., 2014. Knowledge Specialization in Ph.D. Student Groups. IEEE Transactions on Engineering Management 61, 52–67. https://doi.org/10.1109/TEM.2013.2283039

Cornelissen, T., Dustmann, C., Schönberg, U., 2017. Peer Effects in the Workplace. American Economic Review 107, 425–456. https://doi.org/10.1257/aer.20141300

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. Proceedings of the National Academy of Sciences 201422822. https://doi.org/10.1073/pnas.1422822112

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

Falk, A., Ichino, A., 2006. Clean Evidence on Peer Effects. Journal of Labor Economics 24, 39–57. https://doi.org/10.1086/497818

Fox, M.F., 1983. Publication productivity among scientists: A critical review. Social Studies of Science 13, 285–305.

Gaule, P., Piacentini, M., 2018. An advisor like me? Advisor gender and post-graduate careers in science. Research Policy 47, 805–813. https://doi.org/10.1016/j.respol.2018.02.011

Halse, C., Mowbray, S., 2011. The impact of the doctorate. Routledge.

Horta, H., Cattaneo, M., Meoli, M., 2018. PhD funding as a determinant of PhD and career research performance. Studies in Higher Education 43, 542–570. https://doi.org/10.1080/03075079.2016.1185406

Larivière, V., 2012. On the shoulders of students? The contribution of PhD students to the advancement of knowledge. Scientometrics 90, 463–481. https://doi.org/10.1007/s11192-011-0495-6
Lee, S., Bozeman, B., 2005. The Impact of Research Collaboration on Scientific Productivity. Social Studies of Science 35, 673–702. https://doi.org/10.1177/0306312705052359
Lempriere, D.M., 2020. What makes a good PhD supervisor? 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. The American Economic Review 81, 114–132.
Libaers, D.P., 2012. Time Allocation Decisions of Academic Scientists and Their Impact on Technology Commercialization. IEEE Transactions on Engineering Management 59, 705–716. https://doi.org/10.1109/TEM.2012.2184119
Lissoni, F., Mairesse, J., Montobbio, F., Pezzoni, M., 2011. Scientific productivity and academic promotion: a study on French and Italian physicists. Industrial and Corporate Change 20, 253–294.
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
Mangematin, V., 2000. PhD job market: professional trajectories and incentives during the PhD. Research 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. and Pub. Pol. 30, 405–414. https://doi.org/10.3152/147154303781780209
Mas, A., Moretti, E., 2009. Peers at Work. The American Economic Review 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].
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).
OECD, 2012. Closing the gender gap: Act now. OECD Publishing Paris.
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
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. Studies in Higher Education 37, 103–118. https://doi.org/10.1080/03075079.2010.501104
Powell, K., 2016. Does it take too long to publish research? Nature News 530, 148. https://doi.org/10.1038/530148a
Sabatier, M., Carrere, M., Mangematin, V., 2006. Profiles of academic activities and careers: Does gender matter? An analysis based on French life scientist CVs. The Journal of Technology Transfer 31, 311–324.
Shibayama, S., 2019. Sustainable development of science and scientists: Academic training in life science labs. Research Policy 48, 676–692. https://doi.org/10.1016/j.respol.2018.10.030
Stephan, P., 2006. Wrapping it up in a person: The mobility patterns of new PhDs. Innovation policy and the economy 7, 71–98.
Stephan, P., Levin, S., 2002. The importance of implicit contracts in collaborative scientific research, in: Science Bought and Sold: Essays in the Economics of Science. Univ. Chicago Press.

Stephan, P.E., 1996. The economics of science. Journal of Economic literature 34, 1199–1235.
Stephan, P.E., Levin, S.G., 1992. Striking the mother lode in science. the importance of age, place, and time. Oxford.
Tenenbaum, H.R., Crosby, F.J., Gliner, M.D., 2001. Mentoring Relationships in Graduate School. Journal of Vocational Behavior 59, 326–341. https://doi.org/10.1006/jvbe.2001.1804
Waldinger, F., 2012. Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. The Review of Economic Studies 79, 838–861. https://doi.org/10.1093/restud/rdr029
Waldinger, F., 2010. Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany. Journal of Political Economy 118, 787–831. https://doi.org/10.1086/655976
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

APPENDIX 1

Table A.1. Descriptive Statistics of the explanatory variables, by field.

	Engineering				Mathematics				Medicine-biology-chemistry				Physics			
77,143 Ph.D. students	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 programme	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 2

This appendix reports a robustness check in the attribution of the publications to Ph.D. students based on the co-authorship with the supervisor. Specifically, to attribute one paper to a student, we require both the student's name and the supervisor's name to be listed among the paper authors. Using this attribution criterion, we find that 59.79% of the students have at least one paper co-authored with the supervisor during the training period.

Table A.2 shows the descriptive statistics of the newly calculated dependent variables, while Table A.3 shows the regression results.

Table A.2. Descriptive statistics of the students' productivity outcomes. Publication attribution 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 A.3. Regression results. Publication attribution based on the co-authorship with the supervisor. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor characteristics			
Female supervisor	0.028	0.069**	0.37***
Supervisor's seniority	0.11***	0.067***	0.33***
Supervisor's seniority^2	-0.0048***	-0.0032***	-0.015***
Mentorship experience	-0.019***	-0.0081***	-0.041***
Supervisor's publications	0.027***	0.0078***	-0.083***
Supervisor's average citations	0.038***	0.21***	0.23***
Supervisor's co-authors	0.00073***	0.0019***	0.074***
ANR grant	0.16***	0.58***	0.71***
EU grant	-0.15***	0.27***	-0.94***
Team characteristics			
With peers	0.20***	0.23***	0.53***
N. peers	-0.077***	-0.046***	-0.28***
At least one female peer	-0.026	0.071**	0.16
Average peers' seniority	-0.16***	-0.14***	-0.64***
Peers' publications	0.074***	-0.17***	-0.67***
Peers' average citations	0.011***	0.059***	0.067***
Peers' co-authors	0.0011	0.0039	0.17***
Other controls			
French Top-20	-0.063***	0.044	-0.40***
Citation-weighted publications per affiliate	0.013***	0.026***	0.13***
IDEX	-0.021	0.0088	0.035
Department size [100 affiliates]	-0.00042	0.0016**	0.0057**
N. of Ph.D. students in the programme	0.000047***	0.00020***	0.00028***
Female student	-0.36***	-0.18***	-1.05***
Co-supervision	-0.094***	-0.077***	-0.60***
Engineering	0.35***	0.45***	0.92***
Physics	0.72***	0.61***	1.78***
Medicine-biology-chemistry	1.41***	1.44***	5.24***
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	0.25***	-0.25***	1.14***
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.

APPENDIX 3

This appendix reports a robustness check in the attribution of the publications to Ph.D. students based on the similarity between the publication abstract and 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; hence, identifying publications highly similar to the thesis allows us to increase the probability of a correct attribution of the publication to the student. To measure the similarity between a publication and a student's thesis, we rely on a text analysis algorithm comparing the publication and thesis abstracts (Mikolov et al., 2013). We consider only papers with a similarity index greater than 0.8 (the index ranges from -1 to +1). We end up with 44.27% of the students having at least one paper attributed.

Table A.4 shows the descriptive statistics of the newly calculated dependent variables, while Table A.5 shows the regression results.

Table A.4. Descriptive statistics of the students' productivity outcomes. Publication attribution based on similarity.

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 A.5. Regression results. Publication attribution based on similarity. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Supervisor characteristics			
Female supervisor	0.035*	0.087***	0.33***
Supervisor's seniority	-0.015***	-0.033***	-0.085***
Supervisor's seniority^2	0.00075***	0.0012***	0.0042***
Mentorship experience	-0.011***	-0.0028	-0.019**
Supervisor's publications	0.016***	0.0015	-0.078***
Supervisor's average citations	0.025***	0.14***	0.17***
Supervisor's co-authors	0.0013***	0.0016***	0.058***
ANR grant	0.19***	0.69***	0.99***
EU grant	-0.12**	0.15**	-0.85***
Team characteristics			
With peers	0.16***	0.14***	0.53***
N. peers	-0.072***	-0.035***	-0.24***
At least one female peer	-0.012	0.056**	0.16
Average peers' seniority	-0.092***	-0.063***	-0.49***
Peers' publications	0.086***	-0.089***	-0.30***
Peers' average citations	-0.00032	0.028***	0.0048
Peers' co-authors	0.0013	0.0038	0.12***
Other controls			
French Top-20	-0.24***	-0.17***	-1.00***
Citation-weighted publications per affiliate	0.070***	0.073***	0.36***
IDEX	0.050*	0.19***	0.59***
Department size [100 affiliates]	-0.0032***	-0.0038***	-0.012***
N. of Ph.D. students in the programme	-0.00017***	-0.000097***	-0.00054***
Female student	-0.32***	-0.15***	-0.96***
Co-supervision	0.076***	0.044*	0.0055
Engineering	0.19***	0.33***	0.79***
Physics	0.14***	0.034	0.27
Medicine-biology-chemistry	0.69***	0.85***	3.42***
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	1.19***	0.72***	3.66***
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.

APPENDIX 4

This appendix reports a regression exercise where we include a proxy for the student's intrinsic ability among the control variables. Specifically, we identified in our study sample the students who 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)[20]. These contests are organized both at the national and international level, and students showing 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 A.6 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 coherent with previous literature findings (Belavy et al., 2020; Mangematin, 2000).

[20] 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 A.6. Regression results. Including a proxy for the student's ability. OLS estimates.

	(1) Publications	(2) Average citations	(3) Co-authors
Student's ability			
Math Olympiad	0.19	-0.0094	-0.87
Supervisor characteristics			
Female supervisor	-0.0049	0.074**	0.31**
Supervisor's seniority	0.037***	0.0071	0.11***
Supervisor's seniority^2	-0.0019***	-0.00096**	-0.0067***
Mentorship experience	-0.018***	-0.0072***	-0.037***
Supervisor's publications	0.027***	0.0070***	-0.10***
Supervisor's average citations	0.031***	0.20***	0.21***
Supervisor's co-authors	0.0028***	0.0014***	0.091***
ANR grant	0.0050	0.54***	0.22
EU grant	-0.19***	0.33***	-1.28***
Team characteristics			
With peers	0.13***	0.24***	0.25
N. peers	-0.12***	-0.042***	-0.39***
At least one female peer	-0.028	0.073**	0.21*
Average peers' seniority	-0.14***	-0.13***	-0.63***
Peers' publications	0.13***	-0.15***	-0.64***
Peers' average citations	0.0065***	0.056***	0.049***
Peers' co-authors	0.0029	0.0017	0.21***
Other controls			
French Top-20	-0.0084	0.068**	-0.36***
Citation-weighted publications per affiliate	0.012**	0.026***	0.14***
IDEX	-0.056	0.031	-0.031
Department size [100 affiliates]	0.00081	0.0014**	0.013***
N. of Ph.D. students in the programme	0.000092***	0.00023***	0.00038***
Female student	-0.64***	-0.19***	-1.84***
Co-supervision	-0.065***	-0.042	-0.66***
Engineering	0.18***	0.40***	0.99***
Physics	0.77***	0.57***	2.45***
Medicine-biology-chemistry	1.54***	1.39***	6.44***
Mathematics	Ref.	Ref.	Ref.
Entry year dummies	Yes	Yes	Yes
Constant	1.23***	0.23***	3.84***
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.

The UNU-MERIT WORKING Paper Series

2021-01 *Transformation towards sustainale development goals: Role of innovation* by Michiko Iizuka and Gerald Hane

2021-02 *Access to social protection for platform and other non-standard workers: A literature review* by Tamara A. Kool, Giulio Bordon and Franziska Gassmann

2021-03 Labour-augmenting technical change data for alternative elasticities of substitution, growth, slowdown, and distribution dynamics by Thomas Ziesemer

2021-04 *Democracy and COVID-19 Outcomes* by Gokhan Karabulut, Klaus F. Zimmermann, Mehmet Huseyin Bilgin and Asli Cansin Doker

2021-05 *Patent landscaping using 'green' technological trajectories* by Önder Nomaler & Bart Verspagen

2021-06 *Macroeconomic stimulus packages and income inequality in developing countries: Lessons from the 2007-9 Great Recession for the Covid-19 crisis in South Africa* by Alexis Habiyaremye, Peter Jacobs, Olebogeng Molewa and Pelontle Lekomanyane

2021-07 *Deepening or delinking? Innovative capacity and global value chain participation in the ICT sectors* by Rasmus Lema, Carlo Pietrobelli, Roberta Rabellotti and Antonio Vezzani

2021-08 *COVID-19 policy responses, mobility, and food prices: Evidence from local markets in 47 low to middle income countries* by Stephan Dietrich, Valerio Giuffrida, Bruno Martorano and Georg Schmerzeck

2021-09 *Attitudes towards inequality in Brazil: An analysis of a highly unequal country* by Cintia Denise Granja and Ana Maria Carneiro

2021-10 *Mobile phones and HIV testing: Multi-country evidence from sub-Saharan Africa* by Francesco Iacoella and Nyasha Tirivayi

2021-11 *What makes a productive Ph.D. student?* by Alberto Corsini, Michele Pezzoni and Fabiana Visentin