Learning to Play the Game: Student Publishing as an Indicator of Future Scholarly Success

Diogo Pinheiro¹, Julia Melkers¹, Jan Youtie¹,²

¹School of Public Policy, Georgia Institute of Technology, Atlanta, Georgia USA

²Enterprise Innovation Institute, Georgia Institute of Technology, Atlanta, Georgia USA

Corresponding author: Julia Melkers, (email) julia.melkers@pubpolicy.gatech.edu, (voice) (tel) 404-385-0456, (fax) 404-385-0504
Abstract

The need to encourage future generations of students in the pursuit of scientific research has been viewed as a cornerstone of US R&D efforts. However, few indicators of student intellectual activity at the graduate level are tracked on an ongoing basis. The aim of this paper is to examine graduate student publishing as an indicator of pre-doctoral research activity and future scholarly success. This study addresses the gap in knowledge about student publishing through a distinctive dataset that merges bibliometric publication data with survey data from a study of academic scientists. These data are from a nationally representative sample (n=1598) of scientists employed in Research I institutions. For each survey respondent, we have compiled a lifetime publication record from the Web of Science, Science Citation Index. The results indicate that the share of students with at least one publication is substantial and growing over time. Co-publication with advisors is found to be an important driving factor in publication activity, along with certain demographic and field characteristics. Our analysis also suggests that graduate student publication and collaboration is a predictor of later career success and productivity, and as such an important tool in evaluating graduate programs.

Keywords: co-authorship, graduate students, faculty advisor
1. Introduction

There is growing interest in the ability to maintain competitiveness in science, technology, engineering, and mathematics (STEM) fields. Information gathered about STEM activity at the doctoral level mostly focuses on counts of graduates, their demographic characteristics, and whether or not they end up in academic or non-academic jobs. Indicators that address student research activity at the graduate level are rare. Yet publication pressures have become an ongoing feature of the research enterprise, and the ability of graduate programs to foster capabilities for research publication remain in the fore.

This paper’s objective is to examine the role student publishing within the student experience and as a future indicator of career productivity. The paper posits that graduate student publishing rates are rising, associated with certain demographic characteristics and collaborations with the faculty advisor, and indicative of future career article output. The paper combines a national survey of scientists in six STEM disciplines at research universities with the publication record of respondents to the survey excerpted from the Web of Science.

The paper’s findings support hypotheses that student publishing rates are increasing and co-authorship with the faculty advisor makes a difference. Underrepresented groups are observed to be less likely to publish with their advisors, with the role of early socialization evidenced in the positive and significant relationship between students with academic parents and advisor co-publishing. A strong and significant relationship between publication with the faculty advisory prior to receipt of the PhD and productivity in the post PhD career period is evidenced. As such the paper contributes to research on academic publishing by advancing student publishing as an important early indicator of academic career success.

2. Theory

The increasingly competitive academic labor market in the science and engineering fields raises the expectations for doctoral students and post-doctoral candidates as they contend for a limited number of faculty positions, particularly in research universities. Faculty publication rates have been steadily increasing over the past few decades [1], with science and engineering publications showing annual increases of 2.5% from 1995 to 2007 [2]. Additionally, an increasing share of graduate students rely on research assistantships for graduate school funding[2]. These results reflect an increasing norm and expectation in the science and
engineering disciplines for publication, but also the evolving scientific research enterprise as one that is collaborative on a global scale, and organized in large centers, including virtual ones, thereby changing the opportunities one has to collaborate and co-author with other researchers [3]. Mirroring this trend, in today’s competitive academic labor market, the expectation is that doctoral students have a publishing “track record” that demonstrates their technical and scientific expertise, but also that they know how to “play the game” and be productive young faculty members. While academic productivity has become a matter of scrutiny [4], there has also been a growing interest on doctoral student productivity, and participation in this important aspect of academic work.

In ideal terms, the doctoral training process should socialize and prepare students for the increasingly complex world of scientific academic research, and provide a strong foundation from which new faculty may embark on productive and successful early careers. For students in STEM disciplines, a certain proportion of doctoral recipients go on to industry or government research, but many remain attracted to the professorate. In some fields, more than half will hold at least one postdoctoral position before gaining a tenure track faculty appointment, which are also competitive particularly in top institutions. Recent NSF statistics show that the percentage of doctorate recipients that go on to hold a postdoctoral position range from slightly less than 40% for engineering to almost 70% in the life sciences [5].

With regards to post graduation plans, plans to remain in academia range from 14% for engineers to 50% for those in the life sciences. Once hired, the need to publish intensifies, as preparation for interim and tenure review begins. New faculty in STEM fields must not only publish at increasing rates in the preparation for tenure, but also demonstrate grant getting ability. Even within the most competitive institutions, however, there is variation in publication patterns among faculty, with some choosing to exit, and others failing to receive tenure. The question arises about the preparation that young faculty received as doctoral students. How prepared are students for the faculty life in research-extensive academic institutions? Are student rates of publishing mirroring general publication trends to show increasing rates of publications prior to completion of the PhD? Are there demographic differences in these patterns? How do the relationships that form during the doctoral experience matter for
productivity for early career faculty? In particular, how does the relationship with the doctoral advisor, particularly in co-authorship of peer reviewed journal articles coincide with publishing patterns of junior faculty? Are young faculty who have maintained collaborative ties with colleagues from graduate school showing more productive patterns early in their faculty careers? In other words, do those relationships give young faculty a “head start”?

2.1. The Doctoral Training Experience in STEM

While there are a number of issues that affect the ongoing attraction and retention of motivated and productive doctoral students [6-11], the challenges continue through their subsequent entry and success in their academic careers. Graduate student experiences are multidimensional and the doctoral educational process is one that builds technical knowledge in a student’s given field, but also potentially provides important professional socialization to the norms and culture of academic life (Barnes and Austin, 2009). Students develop skills and “know-how” that enable them to not only function within the norms of their discipline, but to also learn the process of writing and submitting publishable research. Studies of student socialization point to the special influence of the faculty advisor in this socialization and skill development [12] but also peers and other individuals that play a role in this developmental process.

Given this, the doctoral educational process provides, at least theoretically, the combination of these skills to enable students to complete their degrees and move forward. In STEM fields, the doctoral experience often is followed by a postdoctoral position (one or more) as a career norm and a stepping stone for faculty positions. From a retention perspective, the role of the doctoral advisor, as well as other experiences of doctoral training, have been pointed to as critical in not only quality training and career preparation, but also in retention and attrition issues, leading to the potential exit from scientific careers [10, 12, 13].

These issues of attraction and retention in doctoral experiences are of course critical in producing a scientific workforce. Yet, what is less understood, is how doctoral training experiences have changed over time, and how they have an effect on faculty career behavior, particularly in that all important pre-tenure/early career stage. We address the student
relationships during that time with advisors and colleagues, and potentially significant effects of this doctoral experience on future career behavior and success.

2.2. Productivity
Participating in research is one of the fundamental duties of doctoral students in STEM disciplines, and publication output is a key indicator of research and knowledge production [14]. An expected outcome of engagement in research, and a potential and important indicator of doctoral student academic socialization may in fact be measured through the student co-authorship/publication activities and outcomes during the pre-PhD time period. Studies of the doctoral student experience show increases in involvement in research, and also the opportunity to co-author papers and make conference presentations [10].

Given the trends that may be observed in the funding of students to engage in research projects, as well as heightened expectations to do so, we expect that the extent of publications authored by students is becoming increasingly more prevalent. This increasing attention to publishing in general is likely to have spillover effects on graduate student activities. More doctorate awardees, particularly in the last decade, might be expected to have published as students than in earlier years. Thus we posit that:

**H1. The rate of graduate student authored publications (prior to receipt of PhD) is increasing over time.**

With attention to the important of early career productivity among scientists, a growing literature has emphasized the importance of mentoring and working with advisors and other faculty as important factors that lead to a productive career and have the potential to counter the effects of disadvantageous background characteristics. Mentoring relationships for both PhD and postdoctoral students has been shown to be invaluable in their professional development, providing both present and later career benefits [15]. Early career mentoring, particularly through informal relationships has also been shown to be beneficial, affecting benefits such as job satisfaction, career support and salary [16]. Hilmer and Hilmer[17], for
example, show that publishing with highly productive advisors explains much of the differences in productivity between graduate students in economics. As Judge et al [8] argue, advisors and other faculty are key in providing students with challenging work and in serving as role models, with such modeling having “both motivational (successful advisors should increase student self-efficacy) and learning (students should acquire more job-relevant skills from successful advisors) effects” (p. 275).

Graduate students are introduced to the conduct of science by their dissertation advisor and relationships between the two groups may range from mentoring to collegial [18]. As such, collaboration with faculty and advisors seem to be an important indicator of program quality and predictor of future student success, on top of (and independent from) the sponsorship effects outlined in part of the existing bibliography on advising [e.g. 19].

Prior research has shown that advisor productivity and reputation has an effect on student productivity [9]. In addition to having important impacts on student careers and productivity, collaboration between graduate students and faculty has been suggested to increase faculty productivity as well [20]. Faculty research productivity may facilitate student publishing by serving in a co-authorship relationship and helping the student through the process of publishing a scholarly article in a learning-by-doing and learning-by-example mode.

Alternatively the faculty mentor may hinder student authorship by for example being too busy to participate in the writing or reviewing of a prospective student-authored scholarly contribution. Studies of faculty mentoring suggest that it is a multidimensional role, involving both instrumental activities – such as research method instruction – and psycho social activities – such as learning how to engage in collaborative behavior [8, 21]. Learning how to write and speak in discipline-specific ways, how to frame research questions, and how to effectively collaborate are all greatly helped by publishing with advisors and supervisors. Recent research has shown this type of research learning to be particularly important in science fields [21], especially as this sort of collaboration and advisor investment is far from uniform or universal [22].

Nonetheless we expect that:
**H2. Student authored publications will be positively affected by collaborations with their dissertation advisor.**

While we recognize that the nuances of the establishment and ongoing interactions between students and advisors is complex, we offer the above hypothesis to address this in a preliminary way. For example, it might be spurious causation in that top students go to top faculty both of whom tend to be productive. Nevertheless, the hypothesis points toward a positive covariation between student authorship and co-authorship with the dissertation advisor.

While the advisor is expected to play a pivotal role in the doctoral socialization process, and resulting publication opportunities, recent work has shown that the experience of doctoral students is variable, particularly regarding the relationship with the advisor and the opportunity to actively collaborate on research [10, 21, 22]. These findings reflect other existing studies that have demonstrated prominent effects of individual and group characteristics (such as discipline) in scholarly publication patterns [23, 24].

At the same time, there are concerns that disciplinary practices are mediated by affiliation with underrepresented groups. For example, studies of doctoral student experiences show that women and underrepresented minority doctoral students are supported less frequently as research assistants than are men and whites [2]. Given this, race and gender may play a role in leading to fewer opportunities for student publishing, and therefore student learning and socialization [25]. Language issues may also present barriers to certain opportunities. Thus, foreign doctoral students who are not native English speakers may inhibit the ability of a student to participate in the writing of research findings as well as communication with English speaking collaborators. The counter proposition is that the high expectations for scholarly output and internationalization of scholarly research may lead to a reduction in the effects of these background characteristics. Given this, we suggest that:

**H3. The rate of student authorship varies based on background characteristics such as race, gender, and citizenship in non-native English speaking countries.**
2.3. Early Career Effects

Studies focused on the doctoral student experience have addressed the multitude of factors that explain productive relationships with faculty advisors. While this is seen as a desirable experience and outcome of doctoral education, particularly in the science and engineering disciplines, it has been largely assumed that this type of early activity makes a difference in early faculty productivity.

The literature on academic careers has identified the relevance of research productivity while in graduate school as a predictor of post graduation productivity [8, 9, 17, 19]. Minor differences in the early career stage matter increasingly as a career progresses because of the effects of accumulation of advantage. Merton [26] has argued that an early start in science attracts greater original contributions to knowledge. Miller, Glick and Cardinal [27] also point to an accumulative advantages model where pre appointment publications help lead precisely to the sort of positions that afford researchers more publication opportunities, pointing to the importance and significance of publications as a graduate student independent of post graduation conditions. A few studies have suggested that one of the best predictors of scientific productivity has been early career or graduate school productivity, something that has been observed across several fields, including Industrial and Organizational Psychology [6, 8], Management and Organizational Science[9], and Biochemistry [28]. However, not all studies agree with this productivity association [24, 29], while some find that organizational contexts rather than early publishing are most important [29].

One caveat is that these productivity studies are small in number and rather dated with respect to STEM disciplines. These mixed findings suggest that there continues to be scope for investigating the nature of graduate student publishing and its association with early career productivity. This proposition will examine but one of the dimensions of outcomes that can accrue from the application of the student publishing indicator to learning evaluation programs. Given this evidence, we posit that:

**H4: Student publishing is positively associated with higher levels of early career publishing in subsequent faculty positions.**
Early publishing as a student could be one important factor behind the “Matthew Effect” in science [30]. The student publishing experience has the potential to be useful in helping explain some of the variation in different career trajectories in academia. The methods and data sources used to test these hypotheses are described in the next section.

3. Methods

Few studies have examined graduate student publishing itself and its outcomes on future publishing on a large scale. The reason for this gap can be explained to some extent by data limitations. Studies of students have been mostly snapshots, with little longitudinal data linking to later career productivity. Moreover, studies focused on publication datasets authored by students are limited in their ability to discern co-authorship relationships. Without degree and hiring dates, fuller analyses of these data are not possible.

The data for this study comes from two linked sources. The first dataset is drawn from a National Science-Foundation-funded 2006-07 national survey of academic scientists and engineers in Research I universities in the United States. The second dataset is comprised of full career publication information, including publication rates and co-authorship ties, for each of the survey respondents.

The survey collected data on individual background, career path, productivity, satisfaction and network data. It was set up to gather data on network content and knowledge exchange but at a national scale. The survey uses an ego-centric network design to explore the respondents’ relationships with the individuals in the respondents collaborative and advice networks, as opposed to the global network of which individuals are members [31]. Through the use of detailed survey questions, respondents describe their networks for select activities and their relations with network members [32]. As a result, the survey captures multiple social and relational dimensions of the collaborative and advice networks that are not accessible through journal publication records or other unobtrusive, secondary datasets. To capture these data, the survey instrument included a series of name generator and name interpreter.
questions. Respondents were first asked to write the names of key collaborators or advisors in research collaboration as well as advice and support networks into five name generator questions. Once the survey respondent provided names in each of the five name generator questions, the respondent was then asked a series of name interpreter questions about each of the individuals they had named. Name interpreter questions addressed the type of the collaboration undertaken with the collaborator, details about the level of relationship and origin of acquaintance, closeness of research expertise, communication frequency, grant activity, and general demographics. Alter-level data (i.e., data about the people in a given respondent’s network, such as people with whom the respondent collaborates) were converted to respondent attribute data through the aggregation of mean or sum values within an individual’s network, depending on desired variable structure. The survey was implemented online using Sawtooth Software®. Individuals were invited to the survey via traditional mail with a series of personalized email follow-ups.

Each of the invitations provided a unique user-id and password and directed the individual to the survey website. The complex nature of the name generator and interpreter questions required a specialized electronic platform. This platform automatically removed duplicate name entries and piped forward the resulting appellations, where they were embedded within the appropriate name interpreter questions. In addition to the social network questions, respondents were asked about their research activities, including grant submission and success rate, teaching and committee responsibilities, attitudes about and involvement in interdisciplinary research, work environment, and detailed demographic and academic background questions. Overall, the survey took between 30 and 45 minutes to complete.

The survey sample of 3,677 was randomly drawn from the population of academic scientists and engineers in six disciplines in Carnegie-designated Research I universities (150 universities.) The selection of the sample was stratified by gender, rank and discipline. The disciplines (biological sciences, chemistry, computer science, earth and atmospheric sciences, electrical engineering, and physics) were selected based on the level of female representation (low, transitioning, and high fields). Taken together, 1598 usable responses were received, for an overall response rate of 44%. Responses were fairly evenly distributed across the six fields,
gender (48% women) and rank (27% assistant professor, 28% associate professor, and 45% full professor.) By citizenship, 20% are non-US citizens (32% of assistant professors, 19% of associate professors and 8% of full professors) are non-US citizens. These proportions are consistent with recent NSF data showing 25% of all physical science, mathematics, computer science, and engineering doctoral instructional faculty were noncitizens, 22% were foreign born citizen, and 53% citizen, born in U.S.[33]

From the name generator data, we obtained the name of each respondent’s PhD advisor, as well as their academic affiliation. Each of these names was verified by project staff using a specific verification protocol developed to ensure that the person exists, and is linked to the respondent. Additional information was then collected for each of the alters including organizational affiliation, rank, gender, and contact information.

For the specific models predicting pre and post PhD publication rates we focus on assistant and associate professors. Our focus on assistant and associate professors has both conceptual and practical reasons. Conceptually, we are interested in the collaborations of early and mid career academics, as that is when we might expect the development of professionally relevant social networks to help shape opportunities for advancement in the respondent’s field. The final sample used in this paper, then, includes 780 assistant and associate professors at Research I institutions. We also retain the full sample and use it in the reporting of overall trends.

The detailed survey data is complemented with lifetime publication data gathered for each survey respondent. These data are important to corroborate the structure of collaborative networks for scientists specific to academic productivity. More specifically, we collected data for 1310 of our survey respondents from the Thomson Reuters Web of Science (WoS). Detailed name identification and disambiguation procedures were employed to identify the publications for each of the survey respondents. Respondents in the physics discipline were excluded from this analysis due to additional complications of cleaning large co-authorship teams (as some of the publications involving physicists had 100 or more co-authors). Overall, the resulting data include 49767 articles, reviews, proceedings papers, notes, and letters. Co-author names were then matched to the collaborative alters named in the survey, again using a detailed and labor-
intensive name disambiguation procedure. An important strength of this data collection and survey-publication data merging is that the productivity of specific dyadic network ties could be determined. The validation of author and co-author names, as a way of verifying the integrity and accuracy of our data, is further discussed in Wang et al [34]. Through this data collection, we were able to gather information on publication level data, which included, besides the name of co authors, the date and place of publication. From the survey data we were able to collect the year the respondents obtained their PhD. With that in hand, we created the following variables:

*Number of publications before PhD*- The total number of publications on record for the period of 6 years before the respondent obtained their PhD, year of PhD inclusive. This might not be sensitive to lag issues, such as publications that were completed before obtaining the degree, but were only published later. In order to deal with this potential issue, we compared our results using a cutoff point of one year after the PhD, but the results were all consistent and similar.

- *Number of publications co-authored with advisor before PhD*. The total number of publications on record for the period of 6 years before the respondent obtained their PhD, year of PhD inclusive, where we identified the advisor as a co-author. For the multivariate models below, we used a binary version of this variable, where 1 indicated any publication at all with the advisor prior to the date of PhD.

- *Number of publications co-authored without advisor before PhD*. The total number of publications on record for the period of 6 years before the respondent obtained their PhD, year of PhD inclusive, where we did not identify the advisor as a co-author. For the multivariate models below, we used a binary version of this variable, where 1 indicated any publication at all with the advisor prior to the date of PhD.

- *Average number of Publications per year after PhD*. We also created a variable that counted the average number of yearly publications by each ego from one year after their PhD until 2010. In the multivariate model, we used the natural log of this variable in order to normalize it.
Besides presenting the data above to test the first two hypotheses, we also estimated a multivariate model to test the third and fourth hypotheses. To test these, we used a treatment effects model, as discussed in Maddala [35], where in the first stage we estimate the likelihood of having co-authored with the advisor, and in the second stage we estimate the impact of that co-authorship has on yearly productivity post PhD. We use this model to be able to parse out the endogeneity of publishing with the advisor prior to PhD, as it is likely that advisors pick the most productive students for co-authorship. In the multivariate models, we restrict our samples to assistant and associate professors to have a more restricted time frame given the increasing opportunities for publication available as careers advance and influence of other factors as a result. For independent variables, we used field and demographic variables, as well as a binary variable that indicates whether the respondent’s parents were academics. We include the latter as an alternative way of measuring socialization into the academic world. We know, from existing research[36], that parental education has a significant impact on students grades and participation in activities at lower educational levels, so here we test the impact of having academic parents in academia on respondent graduate school activity. The results are presented in the next section.

4. Results and Discussion

4.1. Student Publishing Rates

Figure 1 presents the share of respondents that had any publications, regardless of co-authorship, prior to their PhD, by decade of PhD and field. As figure 1 indicates, there has been a near constant increase in the proportion of academics that publish as graduate students across all fields. This finding supports this paper’s first hypothesis.
A shortcoming of our data is that it is based on a survey of current academics, and thus does not include those who graduated and left academia entirely. But given findings that those in academia publish more often than those in industry (Tenopir et al, 1997), our findings would likely be more significant had the sample included everyone. As the sample includes only academics in research I institutions, we are focusing on individuals who are already likely more productive than the average of PhD recipients as a whole.

The figure indicates that just a few decades ago publications by graduate students were a relatively rare phenomenon. Part of the explanation is an ever increasing number of opportunities to publish (National Science Board, 2010), given the greater number of publication venues and journals. As such, we make no distinctions as to the quality and visibility of publication in this article, and recognize that our data does not address these issues. But in a context of ever more publication opportunities, there seems to be higher publication expectations of graduate students. In some fields our data shows that over 90% of recent graduates in some fields currently employed in research 1 institutions published prior to PhD. But what role does the advisor play in these publications? The next section discusses co-authorship patterns of our respondents as graduate students.

4.2. Explaining Student Productivity: Co-Authorship Patterns

Figure 2 presents the share of respondents who co-authored with their PhD advisors prior to their graduation, by field and decade of PhD. The results indicate that co-authoring with one’s advisor has become increasingly common for the fields included here. Over 50% of all respondents in every field to graduate after 2000 have at least one publication co-authored with their advisor, with these rates nearly reaching 70% in some fields, such as chemistry and earth and atmospheric sciences.
This rate of co-authorship is something that, at least for our sample, was not observed at all in four out of five fields prior to 1970. The “sponsorship” model (Reskin, 1979) is increasingly applicable to publishing as well.

Table 1 presents striking results with regards to publication patterns. Not only are publication rates higher for those who co-authored at least once with their advisor, but a majority of their pre-PhD publications were co-authorships with their advisors. Over three fourths of all their publications included their advisor as a co-author. Given the findings in previous studies about the importance of cumulative advantage in an academic career, these results point to the significant role that advisors play in one’s career. The results seem to give some support to hypothesis 2.

We also find that co-authorship patterns vary by demographic characteristics, as Table 2 indicates. Asians were less likely to have published before their PhD than either whites or underrepresented minorities (African Americans, Hispanics and Native Americans), regardless of decade they obtained their PhDs in. These patterns persist if we look at collaboration patterns as graduate students. For most of our sample period, whites are more likely to co-author with their advisors than Asians and underrepresented minorities. These findings may point to some of the barriers that minorities face in academia in establishing social ties and finding successful collaborations [37], and suggest that many of the issues faced by underrepresented minorities within academia may start during the PhD training phase.

Table 3 indicates that males and whites are more likely to have collaborated with their advisors, despite females and minorities having closed the gap for overall productivity in recent
decades. This distinction points to differences in access to collaboration opportunities based on race and gender, supporting hypothesis 3.

[INSERT TABLE 3 NEAR HERE]

Table 4 presents the result of a model predicting whether or not a respondent published with their PhD advisor in the first stage, and its impact on overall career productivity in the second. As mentioned above, this model is reduced to assistant and associate professors:

Model results indicate that recent PhDs are more likely to both publish and publish with their advisors, a finding that is statistically significant at the 0.01 level. Asians and females are less likely to both publish, and publish with their advisors prior to their PhD, at statistically significant levels. The findings are not significant for underrepresented minorities. Reflecting our discussion above about native speakers, we find that US born academics are more likely to publish prior to their PhDs, though that finding is not statistically significant when it comes to publishing with their advisors. To get a sense of socialization issues, we also included a variable concerning whether the respondent had at least one parent who had been a faculty member. We find that there seems to be a significant cultural capital component in publication rates, as those who come from academic families are statistically more likely to publish with their advisors. This may point to the importance of cultural capital as something that persists for longer than existing research has analyzed[36]. Having academic parents might give graduate students a head start in terms of understanding the role of the advisor and in terms of understanding the importance of publication activity early on. These findings support our hypotheses one and three.

4.3. Early Career Effects

To parse out the endogeneity present in a model that tries to predict post PhD productivity based on pre PhD production and collaboration, we used a treatment effects model as described in Maddala (1983). The treatment model presents results that are fully
consistent with the logit model in Table 4. Those with academic parents and more recent PhDs are more likely to have collaborated with their advisors, while Asians and females are statistically less likely to have done so. In the outcome equation, we included two additional variables: whether the respondent in question had a postdoctoral appointment, and whether the respondent in question had published without their advisor prior to PhD. We included these as a way to address causality issues. That is, there is always the possibility that publishing with one’s advisor is the result of the advisor’s selection of more productive students to work with. We try to control for that by doing a two stage model. So it seems that at least part of the mechanism here is that students that collaborate with their advisors learn something, either in the preparation of research on in the submission process, that leads them to become more productive in the long term.

The results in Table 4 show that there seems to be an independent effect of co-authoring with one’s advisor, one that goes beyond the respondent’s productivity. This finding is consistent with the works of Williamson and Cable and Hilmer and Hilmer [9, 17]. We find that those who published at all before their PhD, and those who obtained a postdoctoral position, are more productive in their post PhD years. More precisely, those who at one point held a postdoctoral position have around 11% more publications per year through their careers. This finding suggests that the postdoc period not only offers further training, a norm in some disciplines, and a position when academic openings are not available, but it also provides an opportunity for even more publishing. In addition, those who published at least once without their advisors before their PhD have about 21% more publications, on average, per year. These results are statistically significant at the 0.01 level. More striking, however, is the finding that those who have published with their advisor prior to their PhD have nearly 32% more publications per year post PhD. As these categories (publishing without advisor at least once and publishing with advisor at least once) are not mutually exclusive, the effects are additive. This observation points to a difference between co-authorships with one’s advisor and
publications in general, pre-PhD, and their effect on early career productivity. We cannot make any distinctions here, based on our data, as to whether these differences occur because those who co-author with their advisor end up in positions with more resources or whether it is the effect of the learning and socialization that takes place. Williamson and Cable’s work would seem to support the former while Nettles and Millett’s suggest that latter [9, 10]. But since this sample is restricted to scholars at research one universities, disparities in publication expectations and resources are minimized. As such, the evidence seems to point to an independent, beneficial impact of close collaborations with one’s advisor.

5. Conclusion

Using data for a nationally representative sample of US academic scientists in research one universities, we have found evidence that graduate student publication rates, and rates of collaborations with one’s advisor, have increased significantly over time. As early as 40 years ago, a graduate student publishing a paper before graduation was very rare. Even rarer were collaborations between graduate students and their advisors prior to graduation. Regardless of field, gender and race, up until at least the 1980s fewer than half of our respondents had published as graduate students, and fewer than a third had co-authored publications with their advisors. Things change rather dramatically after that period, and the vast majority of recent graduates have published prior to their PhD, with a majority also having co-authored with their advisors.

We also found that co-authoring with one’s advisor is a significant source of publications for graduate students. Not only do the majority of those who publish as graduate students publish with their advisors, but a majority of their publications are also with their advisors. This result underscores the important role of mentoring in academic preparation, as advisors likely provide both resources and socialization into science. Advisors provide resources such as their
assignment to publishable research projects and contribution to graduate student publications. In addition, advisor participation as a co-author provides students with informal knowledge about how to structure a publication and communicate with editors and reviewers. These results systematically vary by race and gender, but overall there seems to be no doubt that collaborating with one’s advisor is an important step in one’s graduate career.

It is important to outline the limitations of our study. There is a reciprocity issue that is generally present in any attempt to measure the quality of the education provided by advisors and graduate programs. It is “signaling” versus “human capital” perspectives at play, and determining whether the positive effects of any educational intervention are the result of the intervention itself, or the selection process into the intervention is difficult. In our case, it is impossible to state, conclusively, if publishing with one’s advisor leads to greater productivity down the road because of learning and socialization effects, or if it is simply the result of an advisor picking the best and brightest to work with from the start.

We have tried to deal with this potential issue by introducing other measures of individual productivity and socialization in our analyses. That our other socialization measure, academic parent(s), is significant, and that we find an effect of advisor co-authorship on early career productivity net of individual productivity, seems to indicate that a significant component of productivity is tied to what a student learns from his or her advisor by collaborating with them. Because we find an effect of a different measure of socialization in determining collaboration patterns, and because we find an effect of collaboration on productivity that is both larger and independent from student publications without their advisor, there is strong evidence that something else at play besides an advisor’s selection of whom to publish with. Add the variation in collaboration patterns over time, and we can claim that student-advisor collaborations seem to play an important role in setting up cumulative advantage in academic careers. Collaborations with advisors play an important role in determining pre PhD productivity, as well as early career productivity, likely leading to better placement and research resources, as predicted in the existing literature.

Another important limitation of our study is that we do not include any data on publication visibility. As such, we make no claims about quality and impact of publications, or
any differences between graduate student publication with and without one’s advisor. We hope to address this in the future.

Finally, it is important to note the limitations inherent in our sample. As a survey of faculty in Research 1 institutions, we do not have any information on PhDs who are in more teaching intensive environments or even outside academia as a whole. These may present significant issues of sample selection bias, and therefore our results have to be understood in this context. On one hand, such a sample limitation actually presents some benefits. By focusing only on Research 1 institutions, we are focusing only on those academics that are required to produce research. As such, there is less of a need to control for things like preference for teaching intensive environments and the like. On the other hand, we cannot extend our findings to PhDs as a whole, and therefore we are silent as to the impact of advisor collaborations on the productivity and career trajectories of those outside research 1 institutions. In addition, we were not able to perform name disambiguation of physics publication authors because it is not uncommon to see 100 or more authors of a physics publication. Thus these results cannot be generalized to this field without substantial additional resources.

The study has made a contribution to the literature on student publishing, highlighting the importance of the faculty advisor function in graduate student and postdoc publishing experience. It is useful to be reminded that faculty have many duties and multiple affiliations (e.g., departmental, research centers) that can lead to “role strain,” re-prioritization, and less attention given to some of these duties [38]. The findings of this study in terms of the important research collaboration function of faculty with graduate student and post-doctoral advisees suggests that this role deserves further attention in faculty reward systems.

The trends suggest that student publishing will continue its upward trajectory. As a consequence, student publishing can be viewed as an important indicator of social change in the academic community. However, student publications are not necessarily distinguished in academic programs, publicly-funded research, or bibliometric indexes and databases of scholarly articles. Thus, it is not always easy to identify a student publication without doing surveys. One approach to initiating the process of differentiating a student publication is in
federally funded research program reporting. Many federally funded research programs in the US require that recipients of research grants report information on the number and demographic characteristic of graduate students and post-doctorates working on the research funded by these grants. Requirements also stipulate inclusion of a list of articles published under the grant. To these basic reporting characteristics, we suggest that performance measurement and evaluation systems gather additional information on publishing by graduate students and post-doctoral graduates. This information can be added with little extra marginal costs and can serve as an early indicator of future productive careers in STEM research. Without this information, key effects of student publishing might be missed.
Acknowledgements

This work was undertaken with support from the National Science Foundation under award number REESE-#0910191. We deeply appreciate the time that the faculty respondents took in responding to the survey. We also thank our team of graduate students who helped significantly in the cleaning and compilation of the bibliometric data. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the National Science Foundation.
References

Table 1. Student Publication Rates With and Without Advisor Collaboration

<table>
<thead>
<tr>
<th>Published with advisor pre PhD?</th>
<th>Avg Publications with advisor</th>
<th>Avg Publications pre PHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>n</td>
<td>867</td>
<td>867</td>
</tr>
<tr>
<td>Yes</td>
<td>3.09</td>
<td>3.99</td>
</tr>
<tr>
<td>n</td>
<td>443</td>
<td>443</td>
</tr>
<tr>
<td>Total</td>
<td>1,310</td>
<td>1,310</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>1.88</td>
</tr>
</tbody>
</table>

Source: Netwise I data
Table 2: Publishing before PhD

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All Publishing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>6.12%</td>
<td>16.94%</td>
<td>49.33%</td>
<td>67.77%</td>
<td>83.97%</td>
</tr>
<tr>
<td>Asian</td>
<td>0.00%</td>
<td>11.11%</td>
<td>25.00%</td>
<td>54.17%</td>
<td>78.33%</td>
</tr>
<tr>
<td>Underrepresented Minority</td>
<td>40.00%</td>
<td>46.67%</td>
<td>55.56%</td>
<td>88.89%</td>
<td></td>
</tr>
<tr>
<td>Publishing with advisor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>3.85%</td>
<td>8.74%</td>
<td>28.67%</td>
<td>43.80%</td>
<td>63.64%</td>
</tr>
<tr>
<td>Asian</td>
<td>0.00%</td>
<td>0.00%</td>
<td>16.67%</td>
<td>38.36%</td>
<td>55.74%</td>
</tr>
<tr>
<td>URM</td>
<td>20.00%</td>
<td>33.33%</td>
<td>33.33%</td>
<td>55.56%</td>
<td></td>
</tr>
</tbody>
</table>

Source: Netwise I data
Table 3: Publication rates prior to PhD, by gender

<table>
<thead>
<tr>
<th></th>
<th>before 1970</th>
<th>during 70s</th>
<th>during 80s</th>
<th>during 90s</th>
<th>during 2000s</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Publishing Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>10.34%</td>
<td>22.50%</td>
<td>50.77%</td>
<td>70.51%</td>
<td>81.67%</td>
</tr>
<tr>
<td>Female</td>
<td>0.00%</td>
<td>9.30%</td>
<td>43.45%</td>
<td>61.44%</td>
<td>88.04%</td>
</tr>
<tr>
<td>Publishing with Advisor Male</td>
<td>4.65%</td>
<td>10.83%</td>
<td>31.79%</td>
<td>47.44%</td>
<td>61.67%</td>
</tr>
<tr>
<td>Female</td>
<td>0.00%</td>
<td>4.65%</td>
<td>22.62%</td>
<td>36.44%</td>
<td>57.61%</td>
</tr>
</tbody>
</table>

Source: Netwise I data
Table 4: Treatment effects model predicting the impact of Pre PhD collaborations on Post PhD yearly productivity

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Treatment: collaborating with Advisor</th>
<th>Outcome: Log of average yearly publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>0.180</td>
<td>0.137***</td>
</tr>
<tr>
<td></td>
<td>(0.175)</td>
<td>(0.0516)</td>
</tr>
<tr>
<td>Computer Science</td>
<td>-0.664***</td>
<td>-0.0723</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.0637)</td>
</tr>
<tr>
<td>Earth and Atmospheric Sciences</td>
<td>-0.304*</td>
<td>0.0736</td>
</tr>
<tr>
<td></td>
<td>(0.174)</td>
<td>(0.0526)</td>
</tr>
<tr>
<td>Electric Engineering</td>
<td>-0.150</td>
<td>0.0449</td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
<td>(0.0699)</td>
</tr>
<tr>
<td>Asian</td>
<td>-0.313*</td>
<td>0.00966</td>
</tr>
<tr>
<td></td>
<td>(0.175)</td>
<td>(0.0628)</td>
</tr>
<tr>
<td>Underrepresented Minority</td>
<td>-0.00817</td>
<td>-0.188**</td>
</tr>
<tr>
<td></td>
<td>(0.241)</td>
<td>(0.0787)</td>
</tr>
<tr>
<td>Published without Advisor pre PhD</td>
<td></td>
<td>0.209***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0489)</td>
</tr>
<tr>
<td>US Born</td>
<td>0.144</td>
<td>-0.0980**</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
<td>(0.0467)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.334***</td>
<td>-0.105***</td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td>(0.0355)</td>
</tr>
<tr>
<td>Year of PhD</td>
<td>0.0708***</td>
<td>0.0171***</td>
</tr>
<tr>
<td></td>
<td>(0.00943)</td>
<td>(0.00287)</td>
</tr>
<tr>
<td>PostDoc</td>
<td></td>
<td>0.109**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0515)</td>
</tr>
<tr>
<td>Published with Advisor pre PhD</td>
<td></td>
<td>0.318***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.114)</td>
</tr>
<tr>
<td>Academic parents</td>
<td>0.292*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.160)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-141.3***</td>
<td>-33.50***</td>
</tr>
<tr>
<td></td>
<td>(18.81)</td>
<td>(5.676)</td>
</tr>
<tr>
<td>Observations</td>
<td>678</td>
<td>678</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1