What's in a number? How (and why) measuring research productivity in different ways changes the gender gap

Lynn P. Nygaard

Peace Research Institute Oslo (PRIO) &

University College London

lynn@prio.org

Karim Bahgat

Peace Research Institute Oslo (PRIO)

karim.bahgat.norway@gmail.com

Abstract

A persistent finding in studies of research productivity is the 'gender gap', where men seem to publish more academic research than women. However, this gap varies widely from study to study, and little has been done to explore how these claims might be sensitive to what is being measured and how. Using a dataset of publications statistics spanning five years for a Norwegian social science research institute, this paper looks at how (and why) measuring productivity in different ways provides different pictures of the gender gap. Based on the situated context of the institute, we also disaggregate the data by staff category, methodological orientation, and language background, and consider the impact of leaves of absence. We find widely varying measures of the gender gap depending on how we measure and disaggregate, and argue that different bibliometric indicators capture different aspects of research performance, including diversity of output and collaboration, which reflect different publication practices that are both gendered and situated. We suggest that looking at academic writing as a situated - and gendered - social practice offers a potential for deriving more theoretically consistent explanations for both the seeming persistence of the gender gap and the wide contextual variations.

Keywords: gender; academic writing; research productivity; research performance; academic literacies; bibliometrics

Lynn P. Nygaard is a special advisor at the Peace Research Institute Oslo, and doctoral candidate at the Institute of Education, UCL. Author of *Writing for Scholars: A Practical Guide to Making Sense and Being Heard*, her doctoral work focuses on academic writing, gender, productivity, and the path to professorship.

Karim Bahgat is a research assistant at the Peace Research Institute Oslo, working with data analysis, GIS mapping, and data visualization. He holds an M.Phil in Peace and Conflict Transformation from the University of Tromsø, Norway, and focuses on topics related to conflict, political violence, terrorism, and broader socioeconomic issues.

Acknowledgements: This work has been supported by a grant from the Research Council of Norway (Grant number 230705/H20). The authors would like to thank Halvard Buhaug, Jonas Nordkvelle, and Ragnhild Nordås, as well as the editors of this special issue, for valuable comments on earlier drafts.

Introduction

For researchers around the world, excellence in academic writing, or 'research productivity', is usually measured by success in academic publishing, particularly publication in international (English language) scholarly journals. A tremendous amount of research has been carried out on productivity using bibliometric indicators (statistical measures of publications output, citations, or both). One persistent finding over the last 40 years is a gender gap: not only do men seem to produce more publications than women, but men are over-represented among the top producers (so-called research stars, see e.g., White, James, Burke, & Allen, 2012) and women are over-represented among the low or non-producers (those who produce little or no published research) (see, e.g., Creamer, 1998; Kweik, 2015). This is often referred to as the 'productivity puzzle,' based on the observation that the gender gap seems to persist despite increasingly progressive attitudes about women in science (Cole & Zuckerman, 1984).

To be sure, much has changed since Cole and Zukerman's landmark study and Creamer's 1998 review of the literature, and the picture is now far more complex. van Arensbergen, Weijden, and Besselaar (2012), for example, find evidence that the gender gap has disappeared in the younger generation of (social science) researchers – and where any gap exists, women seem to outperform the men. However, the bulk of the current research still suggests that men produce more than women, although there appear to be big differences in this gap depending on context. Geographical location seems to make a substantial difference in how much women produce relative to men (Aiston & Jung, 2015; Frietsch, Haller, Funken-Vrohlings, & Grupp, 2009; Padilla-Gonzalez, Metcalfe, Galaz-Fontes, Fisher, & Snee, 2011). In Norway, where this study takes place, both Bentley (2009) and Rørstad and Aksnes (2015) find academic rank (staff category) to be a stronger predictor of productivity than gender.

Disciplinary differences in the relative performance of men compared to women are evident not only between the natural sciences, humanities, and social sciences, but also within disciplinary subfields (Rørstad & Aksnes, 2015). Within the social sciences (the disciplinary location of the institution in this study) Bird (2011), for example, finds significant differences between social science disciplines in the UK, with women's contributions particularly low in political science (the main discipline represented in this

study). Evans and Bucy (2010) find that women's productivity is much lower than men's in sociology, economics, and political science.

Men and women in the social sciences also seem to differ with respect to concrete publication practices, not just in the amount produced. Using journals as the unit of analysis, Evans and Bucy (2010) find that in the social sciences women publish more qualitative research than men, and men are twice as likely to publish a quantitative article. In the field of international studies, Hancock, Baum, and Breuning (2013) found gendered differences related to research focus (where women were more likely to publish in newer subfields), methodological orientation (men are more likely to publish quantitative research); type of output (men are more likely to author books, but no gender differences in publication of book chapters or non-peer reviewed publications).

These different publication practices raise questions about what exactly is measured in a productivity indicator. Although 'productivity' may sound like a neutral and unambiguous term, it is difficult to both define and measure and there may be a mismatch between the way individual authors conceptualize their own productivity compared to the way in which it is measured in their institutional environment (Nygaard & Bellanova 2018). Outputs that are valued in one context are considered irrelevant in another; textbooks, reports, and popular scientific dissemination are all examples of research outputs that are valued differently in different contexts. Moreover, an indicator is only as reliable as the data that goes into it (see, e.g., Kyvik, 1990; Xie & Shauman, 1998 for a discussion on the difficulty of acquiring high- quality data).; while some research-producing environments regularly collect data on a wide range of outputs, others collect data only sporadically, or rely on the commercial databases (such as Web of Science, Scopus, or Google Scholar) that focus mainly (if not exclusively) on English-language journal articles. Thus, productivity scores are less a simple measure of scholarly activity and more a reflection of which outputs are considered desirable within the context, and more importantly, feasible to measure. For these reasons, most studies on productivity rely on journal article publication as a sole indicator of productivity, although a notable few take into account additional outputs, such as book chapters (e.g., Aiston and Jung 2015; Kyvik (1990)) or patents (Frietsch, et al., 2009).

The question that we address in this paper is whether the size (or even existence) of the gender gap in productivity depends on how productivity is conceptualized and measured – that is, what is counted and how it is counted. Our reasoning is that if writing practices are situated (varying across disciplines, methodological orientations, countries, or institutions), and if academic writing is a gendered social practice where women are concentrated in different demographic groups than men, then the gender gap can be expected to vary depending on the composition of the sample and what is captured by the indicator used (Cameron, Gray, & White, 2013). By analyzing a dataset of publications statistics spanning five years for all researchers within a single Norwegian research institute in the social sciences, we consider how measuring productivity in different ways affects the resulting account of the gender gap.

After presenting a theoretical framework of academic writing as a situated and gendered social practice, we describe the context of the study site, how productivity is conceptualized and measured in Norway, and how this provides the backdrop for our approach to exploring the bibliometric data. Our findings demonstrate some relatively large differences in the reported gender gap depending on the specific indicator used and how the data is disaggregated. Overall, we find that women's measured productivity increased relative to men's when leaves of absence are controlled for, and when the indicators include a wider range of publications (more than just journal articles), fractionalize for co- authorship, and do not add a bonus for publication in high-ranking journals. Within disaggregated groups, however, there were some striking exceptions to this pattern, and the measured gender gap ranged from men producing 80% more than women in one context to women producing 22% more than men in another. We conclude by arguing that in the debate about the productivity puzzle, too little attention has been paid to the gendered and situated nature of academic publication practices and how indicators of research productivity are able – or unable – to capture the complexities of context.

Academic writing as a situated and gendered social practice

The literature on gender gaps in productivity provides a wide variety of explanations for why men seem to produce more than women. Many of these explanations have an essentialist flavour to them: Women by nature prioritize differently, have different preferences, are more perfectionistic, or are more risk- adverse (see, e.g., Kessler, Spector, & Gavin, 2014). These types of explanation take little account of context, of the possibility that being a woman in Japan might be a different experience than being a woman in Norway, that women in a female-dominated discipline in the social sciences might face different expectations than women in a male-dominated STEM discipline (as indicated by Monroe, 2013), or that perfectionism might be a (mal)adaptive response to very real environmental pressures (rather than purely a psychological characteristic) (Sherry, Hewitt, Sherry, Flett, & Graham, 2010). In contrast, the other main set of explanations for why women appear to be less productive than men focuses on structural challenges, such as the cumulative advantage or disadvantage hypotheses (Creamer, 1998; Merton, 1968; Rossiter, 1993), which suggest that relatively small advantages (for men) or disadvantages (for women) compound over time into an increasingly larger gap.

The academic literacies approach, by conceptualising academic writing – including academic writing for publication – as a situated social practice, foregrounds the significance of specific contextual dimensions, including identity, to specific instances of production (Lea and Street 1998; Coffin & Donohue, 2012; Lillis & Scott, 2007; Street, 1995). It acknowledges that individual writers negotiate aspects of identity (what feels meaningful, beliefs about themselves as writers) with expectations from their social setting (the values and common practices present in the various communities to which they belong, including their department or discipline) as they make decisions about what to produce (genre), what constitutes 'good enough' (quality), whether and how to collaborate with other authors (co-authorship), and how to prioritize writing among a myriad of other responsibilities (priority) (Lillis & Curry, 2010; Nygaard, 2017).

Conceptualizing academic writing as a situated social practice makes academic literacies a powerful theoretical tool for exploring how gender might interact with other factors in the production of academic publications. For example, while much productivity research tends to treat institutional context as a background factor that affects all

researchers in that institute the same way, women and men may face different (official and unofficial) expectations for research and publications (Aiston & Jung, 2015; Hancock, et al., 2013). In Norway, where gender equality is among the highest in the world, female academics spend more time advising students, express dilemmas about work-life balance differently than men, and collaborate less internationally (Santiago, Carvalho, & Vabø, 2012; Seierstad & Healy, 2012). Even if they share the same perception of expectations, they may have different perceptions of their own skills (Sherry, et al., 2010). The choice of subject matter might also have a gendered component: Hancock, et al. (2013) find that women are more likely to choose newer subfields and qualitative methodologies, and Knobloch-Westerwick, Glynn, and Huge (2013) find that there is some bias against not only female authors, but also topics that are considered 'feminine' (Knobloch- Westerwick, et al., 2013). When it comes to patterns of co-authorship, women tend to be concentrated in fields with less co-authorship (the humanities and social sciences compared to the STEM sciences) (Hancock, et al., 2013; Santiago, et al., 2012), and even within fields where collaboration is common, women are less likely to benefit from collaborative networks (Bentley, 2011; O'Meara & Stromquist, 2015; Seierstad & Healy, 2012). It is thus not unreasonable to expect that because men and women operate in different contexts within academia – even within the same institute – they might also show different patterns of publication, and score differently on productivity indicators. When the existence of systematically different publication practices is combined with a system that disproportionately rewards some practices more than others (such as the publication of journal articles in English-language journals indexed in the Web of Science), the question of social inequalities arises. An important contribution of the academic literacies perspective is its critical perspective on power; it seeks to understand how writing and publishing practices might reflect social inequalities, as well as to critically challenge and transform these practices (Lea and Street 1998Lillis & Curry, 2010; Lillis & Scott, 2007). This critical perspective on power provides a useful theoretical point of departure for exploring variations in the gender gap that arise depending on how it is measured, and has specifically guided our selection of research questions.

Setting and research questions

The setting of this study is a social science research institute in Norway, with a research staff of about 60 full-time equivalents. Like many other academic institutes in Norway, the research conducted at the institute in this study has an international focus; almost all research output is published in English and research staff are expected to have a high competence in English. The staff itself is mostly Norwegian, but about one quarter come from other countries.

Norway differentiates between the independent institute sector (to which this institute belongs) and the university and college sector: after the Second World War the institute sector was established to focus on applied research, while the university and college sector was expected to focus on more basic research and education. Although the sectors overlap in terms of both personnel and institutional collaboration, their institutional environments differ. The institute sector does not grant degrees and receives only a small core grant from the government, relying on competitive grants for the bulk of its funding. As a result, researchers in the institute sector spend more time writing research grants and less time lecturing and supervising students compared to staff in the universities and colleges. Moreover, much of the research produced is applied in nature and does not result in an 'academic' output, but rather reports tailored to a specific funder. Both sectors, however, are subject to the same measures of productivity used by the Norwegian government to distribute core funding.

In Norway, research productivity is measured (and rewarded) using a complex indicator called the 'Norwegian Publication Indicator (NPI)' (see Aagaard, Bloch, & Schneider, 2015). The indicator accounts for publication type, publication channel, and co-authorship, and is calculated by the Current Research Information System in Norway (CRIStin) database (see www.cristin.no/english/). CRIStin systematically gathers information about all academic journal articles, books, and book chapters published by researchers in all research-producing settings in Norway. To qualify as 'academic' in the CRIStin database, the publication must represent original work targeted at an academic audience (reports, textbooks and popular scientific books are not counted); it must appear in a channel with an ISBN or ISSN and approved routines for peer review (eliminating reports and conference papers); and it must itself be peer reviewed (eliminating, e.g., book

reviews and commentaries). Funds are distributed to institutes based on points awarded for each publication and accredited to the institute through author affiliation.

Publications are first weighted by genre and quality level; that is, articles, books and book chapters are each given a certain number of points, and these points are increased if they are published in a top-tier publication channel (journal or press) (see Table 1).

Discipline-specific committees in the CRIStin board determine which channels are Level 0 (not acceptable), Level 1 (acceptable), or Level 2 (the top 20% in each field).

Table 1: Points awarded to each type of output for both Level 1 (standard) and Level 2 (top-tier) channels in the Norwegian model

	Level 1	**Level 2**
Journal articles	1	3
Book chapters	0.7	1
Books	5	8

Authorship is then fractionalized. Until 2016, authorship was fractionalized simply by dividing the points between the authors. For example, a Level 1 journal article with four co-authors would give each author 0.25 points. However, a new formula was developed to ensure that the natural sciences would receive a greater share of the core funding (Kunnskapsdepartementet, 2015). The new formula fractionalizes the square-root of the number of author fractions (unique combinations of authors and institutional affiliations). For example, a Level 1 journal article with four co-authors (one affiliation each) would give each co-author (and their institute) a total of 0.50 points. Finally, international collaboration is rewarded by multiplying the score by 1.3 when at least one of the co-authors has an affiliation outside of Norway. This would raise the point sum for each author in the previous example to 0.65.

The value of the NPI in the context of this study is precisely that it identifies and assigns a numerical value to different publication practices, making explicit what is often implicit in the evaluation of research performance. The complexity of the indicator allows us to introduce individual elements of this indicator separately (weighting by genre, weighting by level, and fractionalization) so we can observe the impact of each element on the gender gap. Coupling the CRISTiN data with personnel data further allows us to account for leaves of absence and to observe how the gender gap is sensitive to the way the data is disaggregated.

Our five main research questions are as follows:

(1) How does the gender gap change when multiple genres of publication are included and weighted?

Most of the research productivity literature is based on either self-reported data (which can be difficult to quality assure) or commercial databases such as Web of Science or Scopus, which are limited to primarily English-language journal articles. The CRIStin database allows us to access high-quality data on journal articles, books, and book chapters in all languages. This means we can compare measures of productivity based on journal articles only, all three genres without differentiating between them, and all three genres with an additional weighting formula. We can thus explore whether men and women have different publishing patterns with respect to genre, and how this affects their respective productivity scores.

(2) How does the gender gap change when high-prestige publications are given additional weight?

The CRIStin database classifies certain journals and presses as high-prestige under the assumption that publication in these journals can act as a measure of quality, and that such a classification will provide an incentive for scholars to target these journals. By comparing productivity scores generated both with and without the extra weighting for so-called Level 2 publication channels, we can see which groups benefit the most from this weighting, and thus which groups publish more in journals classified as high prestige.

(3) How does the gender gap change when co-authorship is fractionalized and weighted for international collaboration?

The change in the NPI formula to account for co-authorship illustrates not only how difficult it is to construct an indicator that is truly 'field neutral' (Aagaard, et al., 2015), but also how different kinds of collaboration patterns are valued and incentivized by the Norwegian government. The CRIStin data allows us to compare who benefits most when each co-author is given full credit for a publication (which is common for most studies on research productivity and explains why researchers in the natural sciences seem so much more productive than those in other fields), when each author is given a fractionalized share of the credit, and when the new NPI formula is calculated based on the square root of author-shares and an extra bonus for international collaboration.

(4) How does the gender gap change when we disaggregate the data and control for rank, language background, and methodological orientation?

The situated nature of academic writing raises the question of how gender interacts with other group membership, or aspects of identity. We reasoned that the gender gap in productivity at a given institute might look different depending on how the data is (dis)aggregated. For example, many existing studies on the gender gap compare the productivity of all men to all women, but if senior staff are disproportionately more productive than junior staff, and women are concentrated in the junior staff while a disproportionate number of men are found among senior staff, then the measured gender gap might look different if men and women are compared within their staff categories.

For the setting under investigation, we judged three types of disaggregation to be particularly relevant in the context of what CRIStin measures: (i) *Methodological orientation*. We reasoned that within a social science context, those who have a primarily quantitative orientation might co-author more and produce more journal articles, while those with a primarily qualitative (or non-empirical) orientation might lean more towards solo-authorship and publish relatively more books and book chapters. If so, especially if there is also a higher concentration of women with a qualitative orientation (as suggested by Hancock, et al., 2013), then women will be advantaged by an indicator that fractionalizes for co- authorship and disadvantaged by one that does not. (ii) *Language background*: Although the institute has primarily Norwegian staff, 25% of the core staff is from other national contexts and a large share of those who have a small affiliation with the institute (20% position or less) have their primary affiliation in a different country.

Given the geolinguistic imbalance of power in academia that indirectly privileges English-language publications and academics from English-speaking countries (Lillis & Curry, 2010) and the focus on English-language publication at the institute, we reasoned that those researchers who had the bulk of their education in an English-speaking country might feel more confident about writing and publishing in English, and thus appear more productive. We also reasoned that language might have a gender dimension related to academic mobility: that is, that there would be fewer women than men with a non-Norwegian background based at the institute (Bauder, 2015). (iii) *Rank*: The research productivity literature has consistently shown that staff category is a strong, if not the strongest predictor of productivity. We reasoned that the gender gap might be smaller within the same staff category.

(5) How does the gender gap change when leaves of absence are accounted for?

Finally, we reasoned that leaves of absence would have a gendered impact on measured productivity. Norway's parental leave policy allows parents to take up to a one-year leave of absence, to be divided between the two parents, with men taking at least 10 weeks of this leave. Because women generally take the larger portion of this leave, we reasoned that accounting for leaves of absence would benefit women more than it would benefit men.

Materials and methods

We looked at the publications and personnel records of all researchers who had been employed or affiliated (full- or part-time) with the institute for the years 2010 (the earliest year that CRIStin data was available for the institute) to 2014 (the most recent quality-controlled year at the time of data collection) inclusive. The publications dataset is a list of scholarly publications (journal articles, book, and book chapters) credited to the institute in CRIStin. For those authors who did not have all their publications listed in the CRIStin database (e.g. those with an additional affiliation outside the institute) we collected publications information from a variety of open sources including Google Scholar, personal webpages, CVs, and professional networking profiles. In total, our dataset consisted of 91 authors associated with 979 publications during the study period. These publications were then linked to information about the authors: gender, rank, language background, methodological orientation, and leaves of absence (see Table 2). Because we linked publications data to personal information, we informed the involved researchers about the project and gave them the right to withdraw (one male withdrew); ethical approval was granted by the Norwegian Center for Research Data.

Gender and rank were coded based on information from personnel records. For rank, we used the highest achieved academic position for the whole study period. The institute sector in Norway operates with three main staff categories: *junior researcher*, *senior researcher*, and *research professor* and these categories were used for our analysis. The category of *junior researcher* comprises all researchers who have not yet completed a doctoral degree (we excluded research assistants from the sample because they are not expected to produce publications). Most studies on research productivity are unclear about whether doctoral candidates are included; we have explicitly included them here both because doctoral candidates in Norway are expected to produce publications during their fellowship, and because in most social science disciplines in Norway, women make up the majority of doctoral candidates (but remain a distinct minority of professors). *Senior researcher* is the broadest category, comprising all those who have a doctoral degree but who have not yet achieved professorship. This group has perhaps the highest expectation to produce because promotion to a permanent position or professorship depends largely on publications. Finally, the rank of *research professor* is normally achieved through promotion (rather than competition for a chair). An external committee, consisting of at least one member with a primary affiliation in

an institute outside of Norway, evaluates the candidate's contribution to the field primarily by looking at the breadth, depth, and substance of the candidate's publications. The ranks of *research professor* in the institute sector and *professor* in the university sector in Norway have comparable demands for publications (but differ in their expectations for student supervision and grant acquisition). We are aware that these broad categories do not neatly map onto categories of rank used in other contexts. Nevertheless, they allow us to distinguish between three broad staff categories with different expectations for performance in relation to research publication.

Language background was coded based on the language used in the bulk of the individual's education up to and including the doctorate; we obtained information about educational background from CVs, personal webpages and personnel records. When it was unclear, we asked the individual in question. Admittedly, language of educational background is a crude indicator, but it is well-suited for this setting where almost everyone fell into a clear linguistic category of 'Norwegian,' 'English' (individuals from the UK, the US, Canada, and Australia), or 'Other' (individuals from South America, Asia, Africa, or non-Anglo countries in Europe). Because our interest was focused on distinguishing between those from the geolinguistic 'center' compared to those from the periphery or semi-periphery (Lillis & Curry, 2010), we merged the categories 'Norwegian' and 'Other'.

For methodological orientation, we looked at the methodological approach used in the bulk of the author's publications. While many, if not most, of the researchers published occasionally outside their methodological orientation, a primary orientation was generally evident. For simplicity, we coded for either quantitative orientation or qualitative orientation.

For leaves of absence, we re-calculated average productivity for the period based on total time worked rather than five years. We were unable to isolate the reason for the absence, so leaves of absence could cover not only parental leave, but also illness or other reasons. And because information about leaves of absence required access to full personnel records, we had to limit our analysis to only those whose primary affiliation was at the institute, in our case those who were employed in full-time positions for at least 3 out of the 5 years (referred to as the core staff in Table 2).

Table 2: Individual characteristics of the 91 individuals in the data set (core staff in brackets)

		Male n	Female n	Total n
Total		54 [15]	37 [19]	91 [34]
Methodological	Qual	25 [11]	22 [15]	47 [26]
orientation	Quant	29 [4]	15 [4]	44 [8]
Language	English	12 [4]	7 [0]	19 [4]
	Other	42 [11]	30 [19]	72 [30]
Rank	Junior Researcher	10 [3]	14 [6]	28 [9]
	Senior Researcher	22 [4]	17 [10]	39 [14]
	Professor	22 [8]	6 [3]	24 [11]

The final dataset contains data for 91 research staff members who had at least a part-time affiliation with the institute for the five-year period in question, and a near complete list of all their publications for that period. Our dataset does not pretend to be exhaustive or representative of social science research, or even social science in Norway. It is, however, a complete analysis of one institute at a certain point in time, and can thus describe with relative certainty how the gender gap looks within that institute given different measures of productivity and demographic disaggregation. The most serious limitation to the study is the small number of individuals in the sample which prevented both analyses of statistical significance and further disaggregation (for example, looking at gender differences within staff categories within a given methodological orientation). Another limitation was that we did not have reliable data for publications that are not used in the NPI formula (such as popular scientific works, reports, textbooks, edited volumes, and so on) because they are registered irregularly in CRIStin; this prevented a richer exploration of potential gender differences in

how researchers prioritize academic publishing over publications aimed at students or non-academic audiences.

The analysis was based on creating a series of cumulative productivity measures designed to capture average annual productivity for each individual researcher during the five-year period:

1. *Articles only* averages the number of journal articles published during the period to establish a baseline measure. We counted all articles published in journals approved by CRIStin, which means our baseline is likely to be more inclusive than that used by studies relying on the Web of Science or Scopus for their data.

2. *All pubs, unweighted* adds books, and book chapters, but gives all three outputs the same weighting.

3. *All pubs, weighted* applies the Level 1 weighting (as described in Table 1). Researchers who publish relatively more book chapters in anthologies will thus have a reduced productivity compared to the previous measure, and those who publish books will see an increase in productivity.

4. *Level 2 bonus* adds the extra weighting to publications published in Level 2 channels (see Table1). Comparing this measure to the previous one reveals the extent to which men and women publish in higher-ranked channels.

5. *Fractionalized* divides the number of points (after the Level 2 bonus is added) for a given publication by the total number of authors. The more co-authoring that takes place, the greater the drop in productivity compared to unfractionalized measures.

6. The *Norwegian model* is based on the full NPI formula and fractionalizes co-authorship by the square root of the number of authors and multiplies the total number of points by 1.3 when at least one of the co-authors has their primary affiliation outside of Norway, which is called 'international collaboration'. The greater the degree of (international) collaboration, the greater the relative increase in productivity compared to the previous model.

Below we describe the impact that the different measures have on productivity scores for men and women at the institute, both at aggregate and disaggregated levels.

Results

The first set of measures (Figure 1 and Table 3) does not disaggregate the data in any other way than gender, and shows that men's average annual productivity over the five years is about 50% greater than women's when only articles are counted. As soon as other publications are added and weighted, the gap reduces somewhat, suggesting that women have a (slightly) more diverse publications profile than men. When we add Level 2 weighting, the gap increases again, implying that, on average, men publish more than women in top-tier channels. Fractionalizing reduces the gap considerably, but using the Norwegian model increases the gap again. This shows that men co-author more than women, particularly in terms of international collaboration.

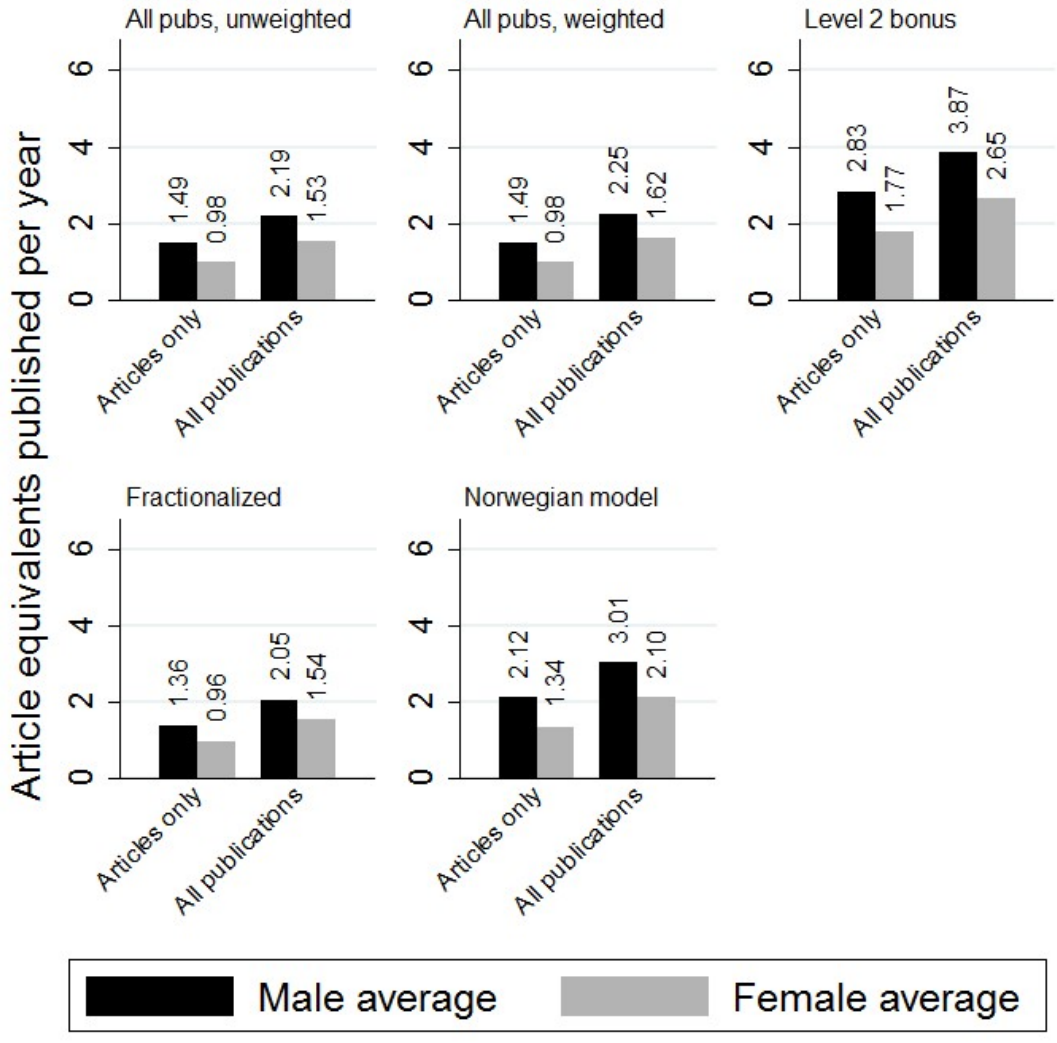

Figure 1: Aggregate measures of average research productivity over five years. Comparison of 'articles only' with 'all publications.'

Table 3: Size of the aggregate gender gap expressed in ratio (men's average productivity relative to women's average productivity).

	Gender gap, Articles only	Gender gap, All publications
All pubs, unweighted	1.52	1.43
All pubs, weighted	1.52	1.38
Level 2 bonus	1.60	1.46
Fractionalized	1.42	1.33
Norwegian model	1.58	1.43

When we disaggregate the data by methodological orientation, some deviations from this pattern emerge. (See Figure 2 and Table 4.) For the quantitative researchers, the gender gap for *Articles only* is a full 80%, dropping down to 62% when other outputs are included. Although the gender gap in this group is much bigger than for the institute as a whole, the general trend is the same: women have a more diverse publications profile. However, a completely different pattern is evident among the qualitative researchers; the gap is only 10% when only journal articles are counted, but when additional publications were added it doubles. This suggests an opposite tendency than the aggregate data: that publishing patterns for women with a qualitative research orientation are *less* diverse than those of men with the same orientation.

The central tendencies are also different when fractionalization is introduced. Although fractionalization reduces the gender gap among the quantitative researchers, it *increases* the gender gap among qualitative researchers. This suggests that among those with a qualitative orientation, women collaborate *more* than the men. When the Norwegian model is introduced, it advantages the women with a qualitative orientation and disadvantages women with a quantitative orientation. This means that the women with a qualitative orientation are not only collaborating more, but also collaborating more internationally than men with the same orientation, whereas the reverse is true for the quantitative researchers. These findings appear to support the notion that women are relatively more productive and have more robust networks in fields that are considered more traditionally 'feminine', such as qualitative research (Knobloch-Westerwick, et al., 2013).

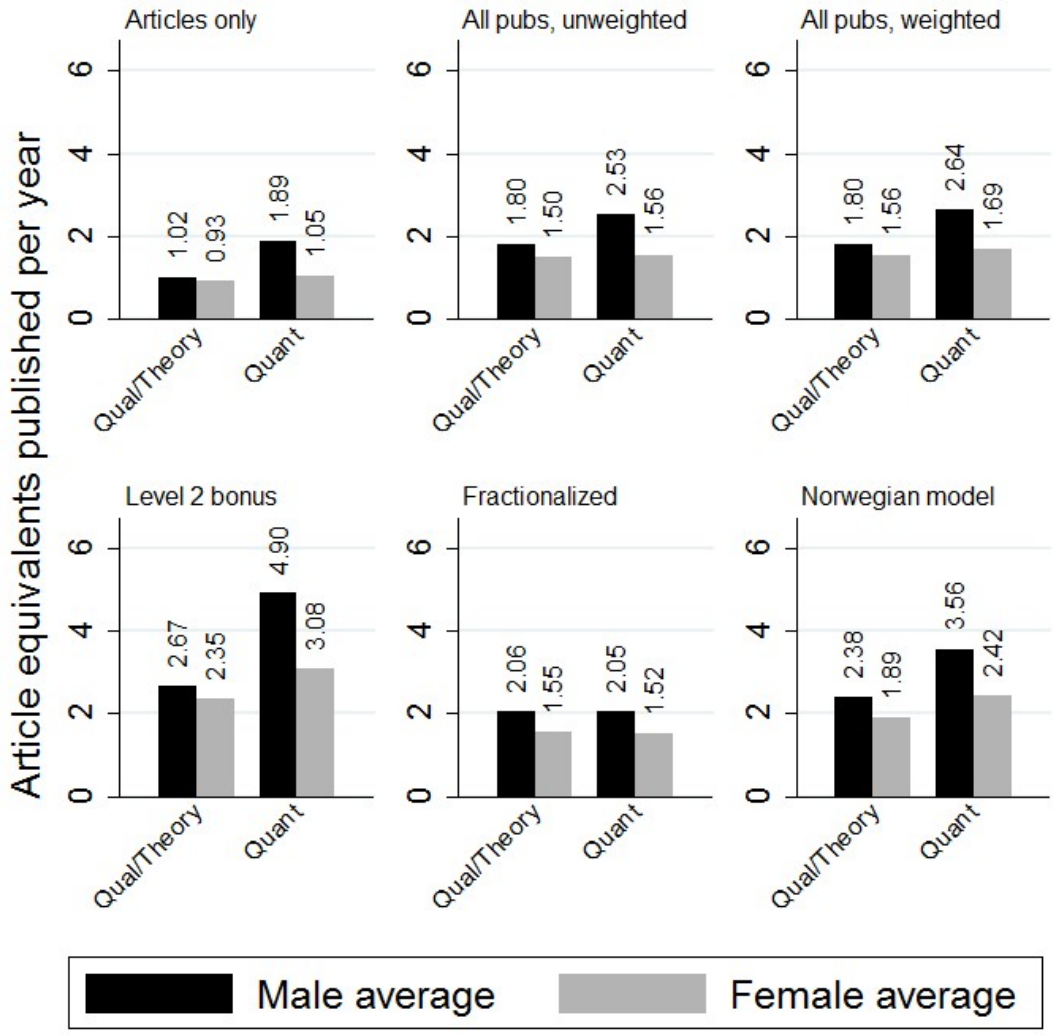

Figure 2: Average research productivity over five years, qualitatively (and non-empirically) oriented researchers compared to quantitatively oriented researchers.

Table 4: Size of the gender gap within methodological orientations expressed in ratio (men's average productivity relative to women's average productivity).

	Gender gap, qualitative researchers	Gender gap, quantitative researchers
Articles only	1.10	1.80
All pubs, unweighted	1.20	1.62
All pubs, weighted	1.15	1.56
Level 2 bonus	1.14	1.59
Fractionalized	1.33	1.35
Norwegian model	1.26	1.47

Next, we turn to language background (see Figure 3 and Table 5). Those with an English language background make up about 20% of the total staff. The most striking finding is that an English language background seems to make almost no difference in productivity for the men, although it does for the women. This raises the question of whether the language background itself makes a difference (perhaps having English as a second language further compounds issues of perfectionism, or a non-Anglo name further compounds gender bias in the evaluation of research), or whether the relationship is spurious, and we are instead seeing something else entirely, such as mobility. Perhaps those who are able to obtain a position in a different country than either their country of origin or the country of their primary affiliation are simply more productive than their non-mobile counterparts. In the specific setting of this study, an English language background is fundamentally related to mobility, and there is a clear gender dimension: the core staff, those who have their primary affiliation at the institute, includes no women with an English language background (Table 2), which is in line with studies showing that women are less mobile than men (Bauder, 2015). All female research staff with an English language background have their primary affiliation at research institutes outside of Norway, and these women are not only more productive than their internal counterparts, but also

more productive than the men as well. This perhaps also says something about how women need to be exceptionally productive to be attractive as a potential collaborative partner (see, e.g., Monroe, 2013), whereas less might be demanded from the men. We cannot draw any conclusions, but the finding illustrates the importance of seeing language and gender not only in context with one another, but also in the context of the specific setting.

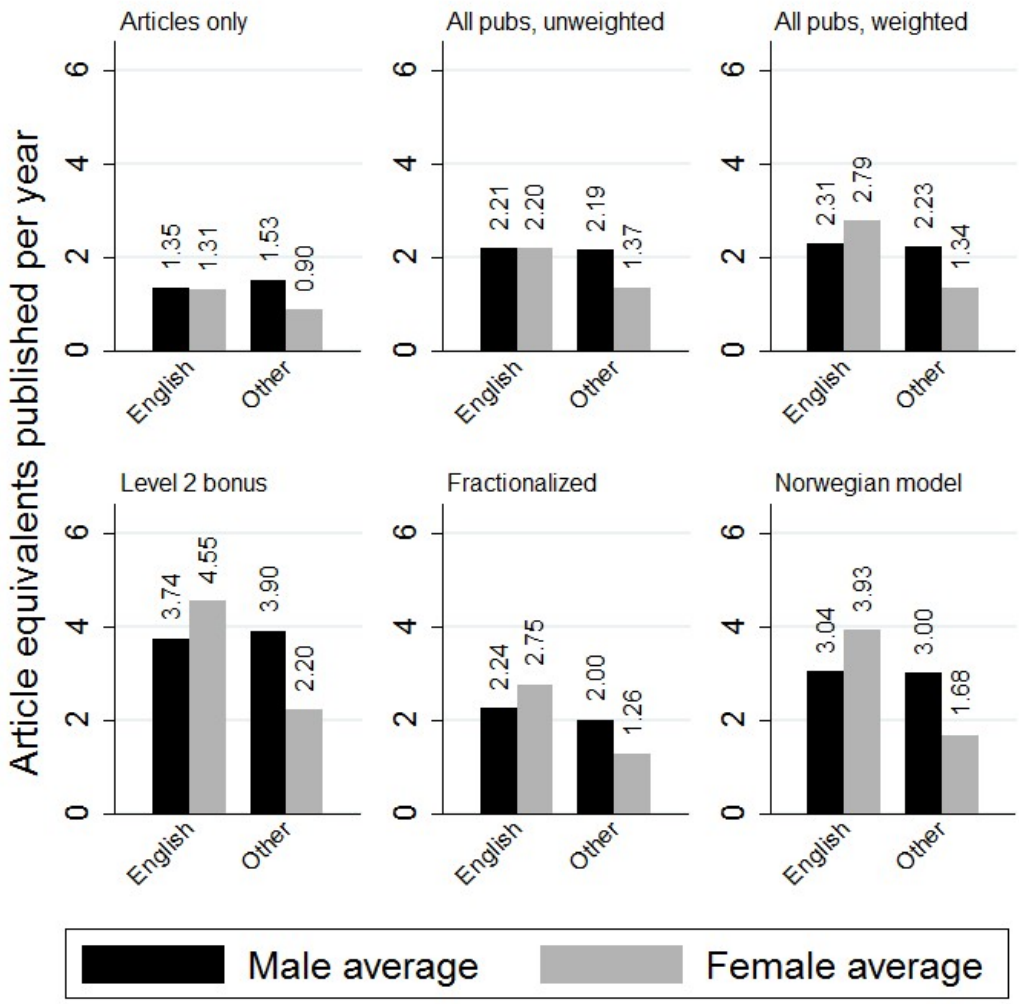

Figure 3: Average research productivity over five years, those with an English language background compared to those with other language backgrounds.

Table 5: Size of the gender gap within language groups expressed in ratio (men's average productivity relative to women's average productivity).

	Gender gap, English	Gender gap, other
Articles only	1.03	1.70
All pubs, unweighted	1.00	1.60
All pubs, weighted	0.83	1.66
Level 2 bonus	0.82	1.77
Fractionalized	0.82	1.59
Norwegian model	0.78	1.79

Looking specifically at those without an English language background, we see a larger gender gap than in the aggregate (suggesting that the presence of a few female research stars among the external women reduces the gender gap for the institute as a whole). The same central tendencies are evident, except that when the outputs are weighted using Level 1 weights, the gender gap increases rather than decreases. This means that even though the women in this group publish more diverse outputs than men, the diversity is mainly in the form of book chapters (since the book chapter weighting reduces overall productivity scores).

We then turn to rank (Figure 4 and Table 6). What is notable here is that, when looking at journal articles only, the size of the gender gap almost doubles each time rank increases (from 12% at the junior level, to 21% at the senior level, and 45% at the professor level). This would appear to support the cumulative advantage hypothesis (Creamer, 1998), where the gender gap gets bigger over time. However, as soon as other outputs are added and weighted, the gap disappears almost entirely for professors, and appears in a reverse direction for junior researchers; only senior researchers still have a modest gender gap. This is, again, a result of women in this study publishing more diverse outputs, especially juniors and professors. For all categories, the addition of Level 2 weighting improves men's relative position to women, suggesting that in all categories, women are publishing slightly less often than men in top-tier publication channels.

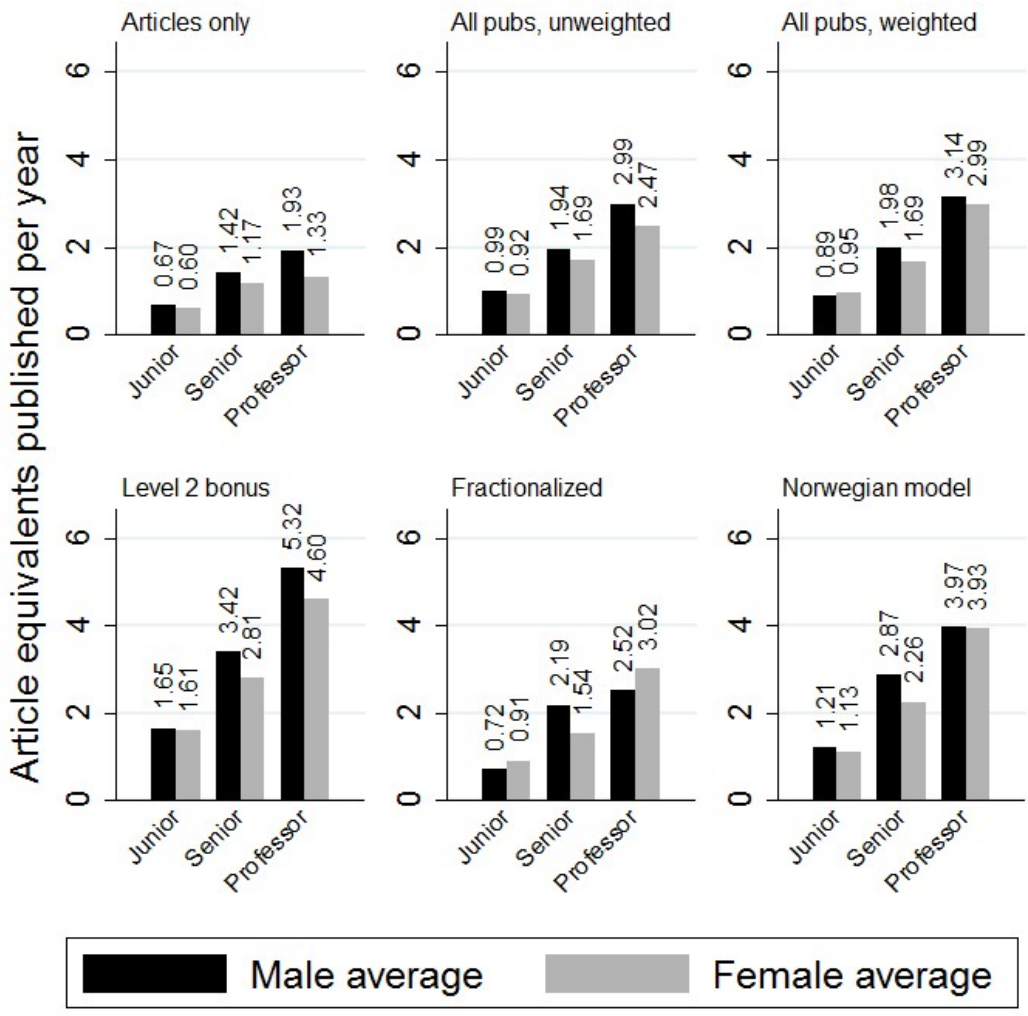

Figure 4: Average research productivity over five years, by rank (junior researchers, senior researchers, and professors).

Table 6: Size of the gender gap within staff categories expressed in ratio (men's average productivity divided by women's average productivity) and percent.

	Gender gap, junior researchers	Gender gap, senior researchers	Gender gap, professors
Articles only	1.12	1.21	1.45
All pubs, unweighted	1.08	1.15	1.21
All pubs, weighted	0.94	1.17	1.05
Level 2 bonus	1.02	1.22	1.16
Fractionalized	0.79	1.42	0.80
Norwegian model	1.07	1.28	1.01

Fractionalization creates an anomaly: for both junior researchers and professors, fractionalization reduces the level of productivity among the men so much that the gender gap is 'reversed' (with women producing about 20% more than the men), which means that the men in this study co-author far more than women at early and late stages in their career. For the middle-ranked senior researchers, however, fractionalizing increases the productivity of the men, implying that the men in this staff category co-author less often than the women.

Because the change in the gender gap at the senior-researcher level moves in the opposite direction compared to that of the junior researchers and professors, it is worth taking a closer look at patterns of productivity between staff categories for each gender separately. Table 7 shows the relative changes in productivity measures as men and women move up the ranks to get a better idea of where the publications practices differ.

Table 7: Ratio of productivity levels between staff categories for each measure.

WOMEN	Jr: Sr	Sr: Prof	MEN	Jr: Sr	Sr: Prof
Articles only	0.51	0.88	**Articles only**	0.47	0.74
All pubs	0.54	0.68	**All pubs**	0.51	0.64
All pubs weighted	0.56	0.56	**All pubs weighted**	0.45	0.66
Level 2 bonus	0.57	0.61	**Level 2 bonus**	0.48	0.64
Fractionalization	0.59	0.51	**Fractionalization**	0.32	0.87
Norwegian model	0.50	0.58	**Norwegian model**	0.42	0.72

Table 7 shows that for both men and women, when only articles are counted, the productivity of juniors is about half that of seniors. In this analysis, we focused on the stability of the ratios: the extent to which a ratio changed within the same staff group as the productivity indicator changed. Of particular interest is how little difference the Level 2 weighting makes for either gender, meaning that the proportion of top-tier publishing changes little in accordance with rank (whereas it might have been expected that the proportion of Level 2 publication would increase substantially as rank increases). For women, what does seem to make a difference is increased output diversity at the professor level: when we look only at articles, seniors are almost 90% as productive as professors. But when we add additional outputs, seniors are only about 70% as productive as professors, and when these outputs are weighted, they are only about 60% as productive as professors.

For men, the most striking impact occurs when fractionalization is added. As described above, when we look at just the change in the gender gap between senior men and senior women when fractionalization is added (Figure 4), it looks as if women have a higher level of co-authorship relative to senior men (because the gender gap increases). However, looking at the ratios from Table 7, what becomes apparent is that the

proportion of co-authorship for women remains relatively stable across staff categories, with fractionalization making almost no difference in the relative gap between juniors and seniors, and only a very slight difference for the gap between seniors and professors. But the men have a much larger change, and in different directions: fractionalizing increases the gap between seniors and juniors and decreases the gap between seniors and professors. This means that male juniors and professors co-author more than seniors. What we might be seeing here is a situation of network building (O'Meara & Stromquist, 2015), where male professors may be more actively mentoring (male) juniors through co-authorship – whereas the senior men might be concentrating somewhat more on solo publication in preparation for promotion to professorship. This suggests that women might be co- authoring less strategically.

The final area we explored in the publications data was the extent to which accounting for leaves of absence could make a difference. We were able to investigate only core staff members, and found that taking into account leaves of absence increases women's productivity by about 12%, compared to a 4% increase for men (see Table 8). This is consistent with our assumption that women in Norway take longer leaves of absence than men, and that absence has an impact on productivity. Importantly, it raises the question of whether productivity should be measured by looking only at output over a fixed period (e.g. five years), or whether the period of time should be adjusted to reflect time worked (e.g., five years minus nine months). The same logic could apply to part-time positions: women make up the majority of those who work part time, and if the total number of hours they are expected to work is not factored into productivity scores, then they will consistently appear less productive than men.

Table 8: Productivity for core staff before and after accounting for leaves of absence

	WOMEN (n=19)			MEN (n=15)		
	Average productivity	**Taking leaves into account**	**Difference (% change)**	**Average productivity**	**Taking leaves into account**	**Difference (% change)**
Articles only	0.80	0.90	+13%	1.36	1.42	+4%
All pubs	1.21	1.37	+14%	2.03	2.12	+4%
All pubs, weighted	1.18	1.33	+13%	2.00	2.08	+4%
Level 2	1.85	2.08	+12%	3.12	3.25	+4%
Fractionalizatiсn	1.23	1.36	+11%	1.67	1.75	+5%
Norwegian model	1.53	1.71	+11%	2.39	2.50	+4%

Discussion and Conclusion

Our aim is not to confirm or deny the existence of a gender gap in productivity, or to claim that the gender gaps reported here would apply directly to other settings. Rather our purpose here is to argue that claims about 'the gender gap' are sensitive to how productivity is conceptualized and measured. Moreover, we suggest that the reason for this sensitivity is that publication practices are highly situated, differing substantively between various groups, and that women are distributed unequally between these groups. If this were not the case, then the differences between men and women's productivity should remain constant regardless of how groups were disaggregated, or how a productivity indicator was constructed. We illustrated this argument by showing how the size or even existence of a gender gap in our study site depended on what was measured (articles only vs inclusion of books and book chapters), how it was measured (how outputs were weighted, co-authorship was fractionalized, and whether leaves of absence were accounted for), and who was included (how we chose to disaggregate).

When we counted only journal articles and compared all women in the institute with all men, it appeared that men produce twice as much as women. However, when we included books and book chapters, the gap started to close. When we gave extra weight to publication in top-tier channels, or gave extra credit for international collaboration, the gap increased again. But when we gave each co- author only their relative share of the credit for each co-authored work, the gap got smaller. Accounting for leaves of absence increased the measured productivity of both men and women, but women substantially more.

Perhaps the most interesting story could be told when we disaggregated the data: in some groups (notably among those with a quantitative orientation, which is traditionally male-dominated), the gender gap appears larger than it does for the aggregate group. In others (notably among the most junior and most senior staff categories), the gender gap appears to have vanished altogether or reversed. Language background seemed to make no difference for men, but it did for women (which made more sense when we considered the issue of mobility). And within some of the disaggregated groups, adding additional outputs or fractionalizing for co-authorship had

the opposite effect than it did for the aggregate group. This suggests that productivity may have less to do with gender than with other group membership (such as rank or discipline), which is in line with findings from other researchers (e.g., Rørstad and Aksnes 2015).

These findings raise questions for how productivity indicators are used in practice to evaluate individual academics, bodies of research, and academic institutes. Using a productivity indicator to say something about an individual academic's performance or make claims about how groups of academics perform relative to other groups presupposes that the activities captured by the indicator are equally relevant for all academics in all contexts. For example, using only journal articles as a productivity indicator to make claims about a gender gap presupposes that both men and women are equally likely to produce a journal article compared to some other output. However, women are more likely to be found in the (qualitative) social sciences and humanities, where books and book chapters are published with greater frequency than in the natural and quantitative social sciences. Likewise, this same disciplinary gender distribution also means that women are less likely to co-author than men, and if authors are given full credit for an article in which they were only one of many authors, then men will appear to be significantly more productive. Similarly, if claims about a gender gap are based on an aggregate measure for an entire institutional population, but women make up the majority of the lowest ranked academics and the minority of the highest ranked, then women will appear to be less productive than if only those of equal rank are compared with one another. In the context of the geolinguistic periphery or semi-periphery (where this study takes place), women tend to be less mobile than men and publish more often in local languages, and may appear less productive than men if only English language publications are counted.

In the discourse on the 'productivity puzzle', researchers need to be conscious about how they conceptualize and measure productivity as well as (dis)aggregate the data. Uncritical use of bibliometric indicators, combined with obliviousness to demographic composition or cultural and institutional context, can result in claims about dramatic differences in productivity between men and women that may well vanish or change when a different indicator is used, or when the data is disaggregated. It is only when we

look at academic writing as a situated – and gendered – social practice that the vastly different results in productivity studies start to make sense and we can start deriving more theoretically consistent explanations for both the seeming persistence of the gender gap and the wide contextual variations.

References

Aagaard, K., Bloch, C., & Schneider, J. W. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation*, 24, 106–117.

Aiston, S. J., & Jung, J. (2015). Women academics and research productivity: An international comparison. *Gender and Education*, 27, 205–220.

Bauder, H. (2015). The international mobility of academics: A labour market perspective. *International Migration*, 53, 83–96.

Bentley, P. J. (2009). *Gender Differences in Research Productivity: A Comparative Analysis of Norway and Australia*. University of Oslo, Oslo.

Bentley, P. J. (2011). Gender differences and factors affecting publication productivity among Australian university academics. *Journal of Sociology*, 48, 85–103.

Bird, K. S. (2011). Do women publish fewer journal articles than men? Sex differences in publication productivity in the social sciences. *British Journal of Sociology of Education*, 32, 921–937.

Cameron, E. Z., Gray, M. E., & White, A. M. (2013). Is publication rate an equal opportunity metric? *Trends in Ecology & Evolution*, 28, 7–8.

Coffin, C., & Donohue, J. P. (2012). Academic Literacies and systemic functional linguistics: How do they relate? *Journal of English for Academic Purposes*, 11, 64–75.

Cole, J. R., & Zuckerman, H. (1984). The productivity puzzle: Persistence and change in patterns of publication of men and women scientists. In M. W. Steinkempt & M. L. Maehr (Eds.), *Advances in Motivation and Achievement* (Vol. 2, pp. 217–258). Greewich: JAI Press.

Creamer, E. G. (1998). Assessing Faculty Publication Productivity: Issues of Equity. In L. J. Scott (Ed.), *ASHE-ERIC Higher Education Report* (Vol. 26, pp. 91). Washington D.C.: The George Washington University, Graduate School of Education and Human Development.

Evans, H. K., & Bucy, E. P. (2010). The representation of women in publication: An analysis of *Political Communication and the International Journal of Press/Politics. PS: Political Science and Politics*, 43, 295–301.

Frietsch, R., Haller, I., Funken-Vrohlings, M., & Grupp, H. (2009). Gender-specific patterns in patenting and publishing. *Research Policy*, 38, 590–599.

Hancock, K. J., Baum, M. A., & Breuning, M. (2013). Women and pre-tenure scholarly productivity in international studies: An investigation into the leaky career pipeline. *International Studies Perspectives*, 14, 507–527.

Kessler, S. R., Spector, P. E., & Gavin, M. B. (2014). A critical look at ourselves: Do male and female professors respond to the same environmental characteristics? *Research in Higher Education*, 55, 351–369.

Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The Matilda Effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, 35, 603–625.

Kunnskapsdepartementet. (2015). Finansiering for kvalitet, mangfold og samspill. Nytt finansieringssystem for universiteter og høyskoler. In (pp. 185). Oslo: Kunnskapsdepartementet.

Kweik, M. (2015). Academic generations and academic work: patterns of attitudes, behaviors, and research productivity of Polish academics after 1989. *Studies in Higher Education*, 40, 1354–1376.

Kyvik, S. (1990). Motherhood and scientific productivity. *Social Studies of Science*, 20, 149–160.

Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157–172.

Lillis, T., & Curry, M. J. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. London: Routledge.

Lillis, T., & Scott, M. (2007). Defining academic literacies research: Issues of epistemology, ideology and strategy. *Journal of Applied Linguistics*, 4, 5–32.

Merton, R. K. (1968). The Matthew Effect in science: The reward and communication systems of science are considered. *Science*, 159(3810), 56–63.

Monroe, K. (2013). Tougher standards for female scholars? The psychology behind them and policies to eliminate them. *International Studies Perspectives*, 14, 476–484.

Nygaard, L. (2007) Publishing and perishing: An academic literacies framework for investigating research productivity. *Studies in Higher Education* 42(3), 519–532

Nygaard, L. P., & Bellanova, R. (2018). Lost in quantification: Scholars and the politics of bibliometrics. In M. J. Curry & T. Lillis (Eds.), *Global Academic Publishing: Policies, Perspectives and Pedagogies*. Bristol: Multilingual Matters.

O'Meara, K. A., & Stromquist, N. P. (2015). Faculty peer networks: Role and relevance in advancing agency and gender equity. *Gender and Education*, 27, 338–358.

Padilla-Gonzalez, L., Metcalfe, A. S., Galaz-Fontes, J. F., Fisher, D., & Snee, I. (2011). Gender gaps in North American research productivity: Examining faculty publication rates in Mexico, Canada, and the U.S. *Compare*, 41, 649–668.

Rossiter, M. W. (1993). The Matthew/Matilda effect in science. *Social Studies of Science*, 23, 325–341.

Rørstad, K., & Aksnes, D. W. (2015). Publication rate expressed by age, gender and academic position – A large-scale analysis of Norwegian academic staff. *Journal of Informetrics*, 9, 317–333.

Santiago, R., Carvalho, T., & Vabø, A. (2012). Personal characteristics, career trajectories and sense of identity among male and female academics in Norway and Portugal. In M.

Vukasović, P. Maassen, M. Nerland, B. Stensaker, R. Pinheiro & A. Vabø (Eds.), *Effects of Higher Education Reforms: Change Dynamics*. Rotterdam: SensePublishers.

Seierstad, C., & Healy, G. (2012). Women's equality in the Scandinavian academy: A distant dream? *Work, Employment, & Society*, 26, 296–313.

Sherry, S. B., Hewitt, P. L., Sherry, D. L., Flett, G. L., & Graham, A. R. (2010). Perfectionism dimensions and research productivity in psychology professors: Implications for understanding the (mal)adaptiveness of perfectionism. *Canadian Journal of Behavioural Science*, 42, 273–283.

Street, B. V. (1995). *Social Literacies: Critical Approaches to Literacy in Development, Ethnography, and Education*. Essex: Longman Group Limited.

van Arensbergen, P., Weijden, I. v. d., & Besselaar, P. v. d. (2012). Gender differences in scientific productivity: A persisting phenomenon? *Scientometrics*, 93, 857–868.

White, C., James, K., Burke, L. A., & Allen, R. S. (2012). What makes a "research star"? Factors influencing the research productivity of business faculty. *International Journal of Productivity and Performance Management*, 61, 584–602.

Xie, Y., & Shauman, K. A. (1998). Sex differences in research productivity: new evidence about an old puzzle. *American Sociological Review*, 63, 847–870.