Gendering excellence through research productivity indicators

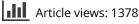## Lynn P. Nygaard, Fredrik N. Piro & Dag W. Aksnes

View supplementary material

Published online: 23 Feb 2022.

Submit your article to this journal

Article views: 1378

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Gendering excellence through research productivity indicators

Lynn P. Nygaard[a], Fredrik N. Piro[b] and Dag W. Aksnes[b]

[a]Peace Research Institute Oslo (PRIO), Oslo, Norway; [b]Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway

ABSTRACT

As the importance of 'excellence' increases in higher education, so too does the importance of indicators to measure research productivity. We examine how such indicators might disproportionately benefit men by analysing extent to which the separate components of the Norwegian Publication Indicator (NPI), a bibliometric model used to distribute performance-based funding to research institutions, might amplify existing gender gaps in productivity. Drawing from Norwegian bibliometric data for 43,500 individuals, we find that each element of the indicator (weighting based on publication type, publication channel, and international collaboration, as well as fractionalization of co-authorship) has a small, but cumulative effect resulting in women on average receiving 10 per cent fewer publication points than men per publication. In other words, we see a gender gap that is not only caused by a difference in the level of production but is also amplified by the value ascribed to each publication.

Introduction

Throughout the world, academics aim for excellence. 'Excellence,' however, is a nebulous concept, difficult to define and measure. The neo-liberal focus on 'new public management' nonetheless pressures academic environments to engage in the measurement and evaluation of excellence at all levels: national, university-wide, departmental, and individual (Mingers and Leydesdorff 2015; Wilsdon 2015). International ranking systems, national performance-based funding systems, and other large-scale evaluation endeavours are often based on the premise that it is important, and possible, to reliably quantify performance and meaningfully compare institutions or individuals. Numerous bibliometric indicators – based on publications, citations, or a combination of the two – have been devised to do just that (Wilsdon 2015).

Critical voices have been quick, however, to point out that these indicators do not always adequately measure what they set out to measure, and that their use sometimes has unintended consequences – such as an emphasis on quantity over quality, or a

Matthew effect, whereby those who have prestige gain additional prestige (Aagaard, Bloch, and Schneider 2015; Furner 2014; Morley 2016; Wilsdon 2015). Nonetheless, their use has exploded in what Gingras (2014) has described as 'evaluation anarchy' over the last decade.

A critical factor behind the explosion of different evaluative bibliometric indicators (and what makes them so difficult to construct) is that writing practices are highly situated and vary across different contexts in academia (Nygaard 2017). What might be considered 'excellent' in one context might be without value in another. For example, collaborating with top-ranking scientists on a report submitted to a United Nations entity might represent a great honour for a researcher, but may not be counted as 'academic output'. Even the production of traditional academic outputs varies from field to field. For example, natural sciences produce journal articles almost exclusively (most of which are authored by large teams), while humanities produce relatively more books and book chapters (with only one or two authors), and the social sciences are somewhere in between (Piro, Aksnes, and Rørstad 2013; Rørstad and Aksnes 2015). This diversity in outputs and practices creates significant challenges for evaluating performance across different contexts. Excellence may be relative, but indicators are relentlessly absolute.

Our question is whether the challenge of quantifying performance and excellence in a context of scientific diversity also has a *gender* dimension. While considerable work has been done to theorize and demonstrate gender bias in general terms in the context of evaluation (Coate and Howson 2016; Lipton 2015; Wilsdon 2015; O'Connor and O'Hagan 2015), to our knowledge no attempt has been made to isolate and quantify the gender impact of specific components of an indicator, or to examine how an indicator might have a different gender effect across different fields. Given the potential of bibliometric indicators to strengthen existing inequalities, and the diversity in publication practices across fields, our aim is to better understand how gender and field might interact to create a gendered Matthew effect.

In this study, we use bibliometric data to analyse the case of the Norwegian Publication Indicator (NPI) as an example of a research productivity indicator based on publications. Unlike the REF in the UK and ERA in Australia, which combine metrics with qualitative peer review, the NPI relies purely on data associated with specific publication practices, allowing us to isolate and measure the gender impact of each component of the indicator across different fields. We do not make claims about gender differences in research productivity per se, but rather demonstrate how indicators of productivity can *disproportionately* benefit men, amplifying whatever gender gaps might already exist.

Theoretical framing: measuring excellence and research productivity in the gendered landscape of academia

Our theoretical assumptions about academic publishing and excellence are grounded in academic literacies theory, which sees academic writing and publishing as highly situated and context-dependent, and academia as a place where power is distributed unequally (Lillis and Scott 2007; Lea and Street 2006). This implies that different disciplines will produce different kinds of writing (Clarence and McKenna 2017), and that some kinds of writing will be given more value than others (for example through evaluation regimes), giving some disciplines or fields of research an unfair advantage.

We thus recognize the challenge of measuring excellence across different environments in higher education and research as formidable. How is it possible to compare a philosopher and a chemist? And in what context will they be compared? 'Excellence' may mean different things depending on whether it is in the context of ranking a university, evaluating the performance of a department within a university, distributing funding, or making hiring decisions. And measuring excellence means making difficult decisions about how to identify the discreet components of excellence and translate them into practices that can be counted (Nygaard and Bellanova 2018; De Bellis 2014).

The element of excellence that we focus on here is research productivity: the extent to which a researcher (or a research-producing institution) produces publications aimed at an academic audience. Identifying quantifiable practices related to research productivity thus requires answering such questions as: Which publications should count? Should some count more than others? How should the credit be divided among authors or institutions?

The answers to these questions matter because any decision about what to include (or not include) in such a metric will, perhaps unintentionally, legitimize some types of output and delegitimize others – thus not only *measuring* productivity, but also defining it and reifying notions of excellence (Moore et al. 2017; Nygaard and Bellanova 2018). For example, if only academic publications are counted, then the production of popular scientific output, or output targeted specifically at stakeholders outside academia, might be seen as less legitimate, less 'excellent'. Metrics can end up favouring specific groups when groups (scientific fields or academic positions) differ systematically in their writing and publication practices. For example, fields that focus on applied research aimed at stakeholders outside the university might be disadvantaged if the metric excludes non-academic outputs. The decisions involved in what to count, and how to count it, then, are directly responsible for conferring prestige (Gingras 2014; De Bellis 2014).

When it comes to gender, bibliometric measures might seem gender-neutral – in that they credit and count publications in the same manner regardless of the author's gender – but they are constructed in an academic landscape that is not. By this we mean that even if women occupy the same academic positions as men, they still navigate a gendered world that places different expectations and demands on them compared to their male colleagues – even within the same academic setting (see, e.g. Witt 2011). For example, women are often expected to take on a larger share of the 'academic housework' (more time in service activities), and often face greater pressure to appear supportive or collegial, making it difficult for them to protect writing time, engage in self-promotion, or decline participation in low-prestige collaborative endeavours (Baker 2010; Lund 2020; O'Connor and O'Hagan 2015; van den Brink and Benschop 2012; Moreley 2016; Coate and Howson 2016). Moreover, women cannot be considered a homogenous group, even in academia; in an academic context, they are professors, lecturers, and doctoral students – and everything in between (see, e.g. Witt 2011). Gender itself is a messy category (with individuals relating to non-binariness in different ways), and it intersects with many other aspects of positionality – including ethnicity, class, and ability, further distributing power unequally throughout the academy. The way 'productivity' is understood and measured might favour high-prestige activities and practices in which (certain groups of) men have greater opportunity to engage.

Women might also lose out from bibliometric measures of productivity simply because of how they are distributed across the academy. In academic settings throughout the world, women are underrepresented in most STEM fields where journal articles are the norm, but far less so in humanities and social sciences where other outputs (such as books and book chapters) are also common (Cameron, Gray, and White 2013; Piro, Aksnes, and Rørstad 2013; Rørstad and Aksnes 2015; Smith 2017). In most fields, women are also underrepresented at the professor level (where most publishing activity takes place) but often make up the majority in entry levels (where the least publishing takes place) (Coate and Howson 2016; Smith 2017).

Uneven gender distribution across the academy, diversity of publication practices in different fields, and the different expectations faced by women all suggest at least three areas of concern in the construction of a research productivity metric: which publication types are included, how high-prestige outputs and activities are accounted for, and how co-authorship is treated. For example, accounting for co-authorship might have a gender dimension simply because researchers in the STEM fields, where women are underrepresented, produce more co-authored works than those in the social sciences and humanities (Aagaard, Bloch, and Schneider 2015; Hug, Ochsner, and Daniel 2014). If all authors are given full credit for every article they take part in writing – regardless of how many other authors they share a by-line with – then purely as a result of demographics, men will appear to be far more productive than women.

Below we describe how research productivity is conceptualized and quantified in the Norwegian Publication Indicator (NPI) and the specific sub-questions we explore.

Empirical focus: the Norwegian Publication Indicator (NPI)

The NPI was developed to help distribute national government funds to research-producing institutions in Norway by rewarding desired publishing behaviour with points that convert to funding, and thus represents a good example of an evaluative bibliometric indicator of research productivity (Sivertsen 2018). While the NPI has been the subject of some controversy in Norway (see, e.g. Sivertsen 2018; Aagaard 2015), it has become well-established and partly copied by some other countries (for example, Denmark, Finland, and Flanders use similar systems, see Aagaard 2018; Pölönen 2018; Engels and Guns 2018).

The NPI draws from a national publication database – Current Research Information System in Norway (CRIStin) – which contains a complete (and quality assured) record of *all* peer-reviewed academic publications produced by researchers in the higher education sector, the independent research institute sector, and the health sector in Norway (Aagaard, Bloch, and Schneider 2015). This is significant because many other settings rely on self-reported data with little quality control, or data from larger independent sources (such as the Web of Science or Scopus) that might be incomplete because they lack data on all publication genres (focusing mostly on journal articles) or publications in languages other than English (Aksnes and Sivertsen 2019).

While many bibliometric indicators are based on citations data (De Bellis 2014; Wilsdon 2016), the NPI is based solely on data associated with the characteristics of the publication. It works by assigning points for each peer-reviewed academic publication using

a highly complex formula (see Sivertsen 2018 for a full description). Here, we focus on the four key components that drive how the points are assigned:

(a) *Fractionalizing co-authorship*: The NPI accounts for co-authorship through 'fractionalization' – that is, by assigning each co-author (or rather, each unique combination of author and affiliation) an equal share of the publication, referred to as 'author shares'. The intent is to make productivity more comparable across fields.

(b) *Rewarding international collaboration:* To encourage international collaboration, the NPI awards publications that include international co-authors (that is, authors with affiliations outside Norway) additional points by multiplying the total point sum by 1.3.

(c) *Weighting for publication type:* The NPI recognizes three types of academic output: journal articles, chapters published in an edited book or anthology, and monographs. If these are published in a recognized, accredited channel with routines for peer review (referred to as 'Level 1 channels'), then journal articles are given 1 point, book chapters 0.7 points, and monographs 5 points.

(d) *Weighting for publication channel:* To encourage publication in top-tier journals and presses, the NPI increases the points in the previous step to 3 points for journal articles, 1 point for book chapters, and 8 points for monographs published in so-called Level 2 channels, which are considered to be in the top 20% of their field. The placement of journals and presses in levels 1 or 2 is determined by national scientific councils in each discipline under the guidance of the Norwegian Association of Higher Education Institutions (see Sivertsen 2018 for more detail).

Since its introduction, the NPI has been extensively debated, both within Norway and internationally. Points of contention include whether it adequately evaluates productivity and allocates funding fairly among the different institutions, whether the method of fractionalization discriminates against large research groups, whether the levels of publication channels should be increased to add further nuance or eliminated entirely, and whether additional publication types should be included (e.g. non-scientific contributions) (Aagaard 2015; Sivertsen 2018). While a negative gender bias has been speculated on in the Norwegian media, it has not been assessed systematically.

Research questions and methodological approach

Because we focus on the extent to which the NPI might amplify already existing gender gaps, we are not concerned with whether men publish more than women, but rather (1) whether the NPI formula disproportionately awards more points *per publication* to men compared to women, and (2) how each specific component of the NPI might drive inequalities. We approach these questions by using the CRIStin data to first analyse differences in average points awarded per publication using the NPI formula, and then for each component of the NPI separately. For both questions, we examine results at the aggregate level and by field.

The study is based on an analysis of 43,500 individuals and their publication output during the 4-year period 2015–2018 (approximately 238,000 publications). We include all individuals with at least one publication during the period analysed; overall, 46 per

cent are women and 54 per cent men. All personnel are included, from PhD students to full professors, as well as medical researchers based in hospitals. The bibliographic data on each individual comes from the CRIStin database, which is coupled to the Norwegian official public registry (NRPR). To determine gender, we use the individual's legal gender recorded in the NRPR, which may differ from the sex assigned at birth.

We disaggregate the analysis by field. CRIStin classifies all publications by field at such a fine-grained level that it would not give us a meaningful basis of comparison. Although there are many possible ways of grouping or subdividing disciplines (and no general solution agreed upon in the literature), for the purpose of our study it is useful to group together subfields which have strong similarities in publication patterns. We thus divide the total population into eight field categories: (1) Humanities, (2) Economics & Management, (3) Social Sciences, (4) Health Sciences (including social medicine, nursery, psychology, etc.), (5) Medicine (including biomedicine and clinical medicine), (6) Natural Sciences, (7) Mathematics & Informatics, and (8) Engineering. We assign each person in our study a field based on where they have the highest number of publications. The fields differ considerably in total numbers of people included, as well as with respect to the distribution of men and women. (See Table S1 in the Supplemental Online Material for detail.) Men and women are also distributed unevenly across fields; we thus weight the result for each gender according to the overall distribution of individuals across fields in our calculation. The weighted averages appear in the Totals in all tables (except Table 3).

We use an observation period of four years to provide larger and more robust datasets for each person. As we lack data on employment status of the individuals, we have not been able to adjust for career length, and some people have not been employed throughout the entire period. We assume, however, that this source of error would affect both genders equally. Similarly, data on leaves of absence are not available, and given the long period of parental leave in Norway this factor would likely affect women more than men. Although the sources of error from not accounting for career length and leaves of absence are relevant to the total number of publications produced, they are less relevant for the total points ascribed to each publication, which is our focus here. For each component of the indicator, we calculate averages for both genders by field, as well as a female/male ratio. A ratio of 1.00 represents gender parity: A ratio below 1.00 means that women score lower than men, and a ratio above 1.00 means that women score higher. The more the ratio differs from 1.00 (either higher or lower), the bigger the gender gap.

The unit of analysis is the individual researcher, which means that each individual counts as one unit regardless of how many publications they have produced. This avoids the skewness in the data created by the common problem of having a small number of prolific researchers producing most of the publications (Kyvik 1991, 90). It should be noted that we do not apply tests of statistical significance in the analyses because the study is based on the entire population of Norwegian researchers rather than a sample.

Results

To establish a baseline, and answer our first research question, we present the average number of publications produced by men and women in each field, the average NPI

Table 1. *Baseline*: Field and gender differences in average number of publications per person (whole counts), NPI points per person, and NPI points per publication.

Field	Publications per person			NPI points per person			NPI points per publication		
	M	W	Ratio W/M	M	W	Ratio W/M	M	W	Ratio W/M
Humanities	4.1	3.5	0.85	5.0	3.7	0.74	1.23	1.06	0.86
Economics & Management	4.5	3.1	0.69	3.8	2.3	0.61	0.85	0.71	0.84
Social Sciences	4.8	3.6	0.76	4.2	3.0	0.72	0.85	0.80	0.93
Health Sciences	6.3	4.1	0.65	3.9	2.3	0.60	0.56	0.54	0.96
Medicine	7.1	4.6	0.65	3.4	1.9	0.56	0.40	0.36	0.90
Natural Sciences	8.6	4.8	0.56	3.4	1.8	0.52	0.50	0.42	0.85
Mathematics & Informatics	5.1	3.3	0.65	3.9	2.2	0.56	0.76	0.64	0.84
Engineering	6.1	4.4	0.72	3.3	2.2	0.68	0.50	0.48	0.96
Total (weighted)	6.5	4.3	0.66	3.7	2.3	0.61	0.60	0.54	0.90

score per person, and the average publication points per article (Table 1). The first set of columns show total production per person, with notable variation between fields. Converting the outputs to points using the NPI increases the gender gap in all fields: both in average publication points per person and in average publication points per publication. In other words, the NPI appears to amplify the gender gap in productivity by disproportionately giving women fewer points per publication than men – on average by 10 per cent.

The questions for the remainder of this paper are to see how each individual component of the NPI might work to amplify (or mitigate) the gender gap in how each publication is valued, and how this might play out differently across fields.

Fractionalizing co-authorship

Fractionalizing co-authorship means that each co-author receives a fraction of the total points rather than counting a co-authored publication as a 'whole' publication in the same way as a solo-authored piece. Table 2 illustrates field and gender differences in patterns of co-authorship and the impact that fractionalization has on the baseline scores.

We see large differences across fields. In the Natural Sciences and Medicine, the average author collaborates with 8 authors per publication (and just 1 per cent of publications are solo-authored), while in the Humanities there are seldom more than two

Table 2. *Fractionalizing co-authorship*: Field and gender differences in co-authorship practices, and impact of fractionalizing co-authorship on points per publication in the NPI.

Field	Average number of authors per publication			Percentage of solo- authored publications (field average)	Impact of NPI	
	Men	Women	Ratio W/M		Men	Women
Humanities	1.8	2.0	1.10	60%	- 23%	- 26%
Economics & Management	3.0	3.1	1.04	14%	- 57%	- 58%
Social Sciences	2.5	2.5	1.01	30%	- 43%	−44%
Health Sciences	5.2	5.0	0.96	5%	- 71%	- 71%
Medicine	9.1	8.6	0.94	1%	- 84%	- 84%
Natural Sciences	8.7	8.6	0.99	1%	- 80%	- 82%
Mathematics & Informatics	3.3	3.4	1.02	8%	- 62%	- 63%
Engineering	4.4	4.7	1.06	3%	- 72%	- 73%
Total (weighted)	6.1	6.0	0.98	8%	- 69%	- 70%

Table 3. Illustration of interaction between fractionalization calculations and gendered patterns of collaboration.

John	Number of authors	Fractionalized co-authorship	Jennifer	Number of authors	Fractionalized co-authorship
Publication A	1	1.00	Publication C	5	0.20
Publication B	10	0.10	Publication D	6	0.17
Per publication	2.75	0.28	Per publication	2.75	0.09

authors per publication. Indeed, 60% of the publications in the Humanities have only one author. We note only minor gender differences, however. In some fields (mostly those where co-authoring is the norm), men have slightly higher numbers of co-authors than women, while in other fields (specifically Humanities, Economics & Management, and Engineering) women co-author slightly more than men. Overall (all fields combined), the gender difference is almost zero.

The impact of fractionalizing co-authorship has a large impact on the total number of points attributed each publication (see Table S2 in the Supplemental Online Material for more detail). Overall, fractionalization reduces the number of points per publication – for both men and women – by about 70%. However, there are large differences across fields that tend to correspond inversely with the first column: For example, in the Humanities, the reduction in points per publication is larger for women than for men, which can be explained by women co-authoring relatively more than men.

One unexpected finding is that women seem to lose more than men do from fractionalization, even in fields (such as Health Sciences, Medicine, and Natural Sciences) where they co-author less. Logically, we would expect fractionalization to always benefit those who co-author less. This surprising result is an artefact of how fractionalization is calculated and gender differences in collaboration patterns whereby women are both less likely to solo-author an article and less likely to work in large teams. A hypothetical illustration of this phenomenon is depicted in Table 3: Here, John and Jennifer have both published two articles. John has one article as the sole author and one with 10 authors. Jennifer has published one article with 5 authors and one with 6 authors. Although they have the same average number of co-authors per publication, the fractionalized value is different. This phenomenon means that in our dataset we see that, in most of the fields, the fractionalization method surprisingly benefits men more than women, but generally only by one percentage point.

Rewarding international collaboration

The NPI increases the total number of points to be divided between authors by 30% when at least one of the authors has an international affiliation. Table 4 shows some striking gender differences in international collaboration in most fields. In percentage points, this difference is highest in Mathematics & Informatics (9 percentage points) and Health Sciences (6 percentage points). Measured as ratio, it is highest in Social Sciences (26 percentage points). When seen in combination with Table 2, this suggests that women have relatively more domestic co-authors than men, while men have more co-authors from abroad. Thus, the extra weight given to internationally co-authored publications could be assumed to benefit men more than women.

Table 4. *International collaboration bonus*: Field and gender differences in average proportion of publications with international co-authorship (based on whole counts) per individual, and impact of international collaboration bonus on points per publication in the NPI.

	Publications with international co-authorship			Impact of NPI	
Field	Men	Women	Ratio W/M	Men	Women
Humanities	12%	12%	1.00	+ 1.6%	+ 1.6%
Economics & Management	37%	33%	0.89	+ 7.5%	+ 6.9%
Social Sciences	20%	15%	0.74	+ 3.1%	+ 2.3%
Health Sciences	36%	30%	0.84	+ 7.0%	+ 5.9%
Medicine	44%	40%	0.92	+ 9.6%	+ 9.0%
Natural Sciences	59%	55%	0.93	+ 13.1%	+ 12.7%
Mathematics & Informatics	40%	31%	0.79	+ 8.7%	+ 6.7%
Engineering	32%	30%	0.94	+ 6.7%	+ 6.2%
Total (weighted)	39%	35%	0.90	+ 6.3%	+ 5.7%

However, Table 4 also shows that the gender differences in points per publication caused by introducing the bonus for international collaboration is minor in every field (see also detailed statistics in Table S3 in the Supplemental Online Material). This suggests that while men collaborate internationally more than women do, the overall proportion of international collaboration is too small to make any substantial impact on the gender gap. Also, the bonus (30% extra weight) is too small to result in large differences overall. Nevertheless, we observe that in all fields except the Humanities, men benefit slightly more than women do by this component of the NPI.

Weighting for publication type

The NPI weights journal articles, book chapters and monographs differently. An amplified gender gap could be driven by, e.g. women producing more low-valued outputs (i.e. book chapters) or men producing more high-valued outputs (i.e. monographs). Table 5 shows gender differences in the production of publication types across fields. For example, the aggregate figures in the bottom row show that 87% of the publications men produce are journal articles, 12.3% are book chapters, and 1.0% are books, compared to 88%, 11.8% and 0.7%, respectively, for women.

Table 5. *Weighting for publication type*: Field and gender differences in average proportions of journal articles, book chapters and monographs (based on whole counts) per individual, and impact of weighting publication type on points per publication in the NPI.

	Journal articles			Book chapters			Monographs			Impact of NPI	
Field	M	W	Ratio W/M	M	W	Ratio W/M	M	W	Ratio W/M	M	W
Humanities	56%	60%	1.08	37.5%	36.3%	0.97	6.6%	3.9%	0.59	+ 16.5%	+ 4.6%
Economics & Management	84%	80%	0.95	13.8%	18.7%	1.35	1.8%	0.8%	0.46	+ 4.9%	−1.1%
Social Sciences	64%	65%	1.01	32.8%	32.8%	1.00	3.1%	2.1%	0.69	+ 3.2%	−1.3%
Health Sciences	93%	94%	1.01	6.7%	5.9%	0.88	0.5%	0.1%	0.23	+ 2.4%	−1.9%
Medicine	99%	99%	1.00	0.6%	0.9%	1.56	0.0%	0.0%	0.81	−0.1%	−0.2%
Natural Sciences	96%	97%	1.01	3.8%	3.1%	0.83	0.1%	0.0%	0.26	−1.0%	−1.3%
Mathematics & Informatics	82%	82%	1.00	17.5%	17.2%	0.98	0.3%	0.9%	2.63	−3.1%	−0.1%
Engineering	82%	85%	1.04	17.9%	14.6%	0.81	0.1%	0.1%	0.98	−5.0%	−4.7%
Total (weighted)	87%	88%	1.01	12.3%	11.8%	0.95	1.0%	0.7%	0.62	+3.2%	−0.6%

As expected, there are some marked differences between fields in the production of journal articles, with Humanities and Social Sciences standing out as the fields that produce the smallest share compared to other publication types. However, there are almost no gender differences within each field.

Relative shares of book chapters and monographs show much greater variation, not only between fields, but also between genders. But for the NPI weighting to have any real effect on a gender gap, the publication type must comprise a reasonable proportion of the total output and there needs to be a considerable gender difference in the production of that publication type. The column showing the impact of the NPI in Table 5 shows that there are notable differences across fields and gender (see also detailed statistics in Table S4 in the Supplemental Online Material). As expected, the number of points per publication is lower in fields with larger proportions of book chapters (e.g. Engineering), while the opposite holds for fields with many monographs (e.g. Humanities). In most fields, men tend to benefit more from the weighting of publication types than women do because, although men publish slightly more book chapters than women do in most fields (which would lower their average scores), they also generally publish more monographs. As monographs have a much greater weight (one monograph corresponds to 7–8 book chapters), the differences in monograph publishing make up for the slightly larger volume of book chapters. Overall, we observe that weighting for publication type increases the number of points per publication for men by 3.2%, while for women there is a reduction of 0.6%.

Weighting for publication channel

To incentivize publication in high-prestige channels, the NPI gives additional points for publishing in top-tier (Level 2) journals or presses. An amplified gender gap could be driven by men publishing more often in Level 2 journals than women do. Table 6 shows that, indeed, in all fields except Engineering, men have a higher share of their publications (based on number of author shares) in Level 2 outlets. However, except for Mathematics & Informatics and Engineering, gender differences appear minimal.

The impact column in Table 6 shows that the overall effect of weighting for publication channel is quite high (much higher than the impact of weighting for international collaboration or publication type). (See Table S5 in the Supplemental Online Material for

Table 6. *Weighting for publication channel.* Field and gender differences in average proportion of Level 2 publications (based on author shares) per individual, and impact of weighting for publication channel on points per publication in the NPI.

Field	Share of publications in Level 2 channels			Impact of NPI	
	Men	Women	Ratio W/M	Men	Women
Humanities	23%	22%	0.94	+ 26%	+ 25%
Economics & Management	22%	21%	0.94	+ 34%	+ 27%
Social Sciences	20%	19%	0.94	+ 23%	+ 24%
Health Sciences	20%	19%	0.97	+ 28%	+ 31%
Medicine	19%	18%	0.96	+ 32%	+ 31%
Natural Sciences	23%	22%	0.92	+ 44%	+ 39%
Mathematics & Informatics	22%	18%	0.81	+ 44%	+ 30%
Engineering	18%	21%	1.16	+ 33%	+ 36%
Total (weighted)	20%	20%	0.97	+ 31%	+ 30%

detailed statistics.) The high impact of Level 2 weighting is related to the fact that an article published in a Level 2 journal is awarded three times more points than one published in a Level 1 journal. Level 2 weighting means that the overall average of number of points per publication for both genders increases by approximately 30%.

However, except for Mathematics & Informatics, men and women seem to gain almost equally much from the extra weighting of Level 2 publications. While in most fields, men gain slightly more than women, there are three fields (Health Sciences, Social sciences, and Engineering) where women gain slightly more than men.

In sum, the individual components of the NPI (the different kinds of weighting and fractionalization) each make minor contributions to amplifying the gender gap. We noted that the overall difference is on average 10 per cent fewer publication points per publication for women compared to men (Table 1), which breaks down to 0.06 points per publication. Looking at each component separately, we see that – of these 0.06 points – 0.01 points (17%) can be attributed to fractionalization, 0.01 points (17%) to rewarding international collaboration, 0.02 points (33%) to weighting of publication types, and 0.02 points (33%) to weighting of publication channels. On the one hand, these figures are arguably quite small. On the other hand, as the figures are calculated per individual publication, these relatively small differences can have a large impact on the aggregate numbers.

Discussion and conclusion

Indicators of productivity are not simply neutral counting systems: They institutionalize what matters, what is valued, and what gets rewarded. And sometimes they can systematically benefit some groups more than others and reinforce inequalities (O'Connor and O'Hagan 2015; Moreley 2016). The purpose of this article has been to demonstrate how different elements in a research productivity indicator can be gendered.

As a point of departure, our data shows (with some notable exceptions that we have pointed out above), very little difference between the genders when it comes to publication patterns within fields. The biggest differences are between fields, not between genders. Yet – despite this minimal difference in publication patterns – the way the NPI counts and awards points results in women obtaining on average 10 per cent fewer publication points *per publication* compared to men, which we describe as an amplified gender gap. In other words, we see that the value ascribed to each publication disproportionately benefits men.

We further looked specifically at how each individual component of the NPI might drive this amplified gender gap, and how the impact might vary across fields. At an aggregate level, none of the individual components seem to have a particularly strong impact on the amplified gender gap. Rather, each individual component has a relatively small, but cumulative effect. However, we see a different story emerge when we look at the field level. Even without isolating each component, we see from Table 1 that the total amplified gender gap varies from less than 5 per cent for Engineering and Health Sciences to 16 per cent for Economics & Management and Mathematics & Informatics. When we look at the impact of each component individually, the differences between the fields become even more evident. Instead of just a small, but cumulative overall effect for each of the NPI components, we see that each of the components has a

disproportionately strong impact on at least one of the fields. For example, in the Humanities, more monographs by men contribute significantly to the gender difference, while in Mathematics & Informatics a higher share of internationally co-authored publications and Level 2 publications by men are main drivers of the amplified gender effect.

This suggests that despite general similarities between men and women in publication patterns, some avenues of excellence appear to be more gendered than others. As pointed out by Lund (2020), gendered divisions of labour both inside and outside academia mean that men are better positioned to produce what counts. The publication of monographs stands out in this respect. While this might be another aspect of the compositional effect (where writing a monograph is associated with a higher-ranked position, where men are better represented), it might also be related to social expectations for women's labour in the home, giving men more freedom to pursue a sustained writing project outside of work hours (Vabø et al. 2012). A consistent finding concerning gender and publishing is that men seem to produce more than women (which can also be seen in Table 1, where at the aggregate level the men in our study produce about 34% more than the women), and recent studies indicate that women also have been more negatively affected by the pandemic in terms of publication productivity (Squazzoni et al. 2021). Likewise, women's reduced engagement in international collaboration might be a compositional effect (where women are concentrated in sub-fields such as education or social work where international collaboration is irrelevant if the subject of study is domestic practices), or it might be a result of women not being as mobile as men and unable to develop international networks.

But what does this amplified gender gap mean for women in practice? The NPI is not intended to be used to make assertions about productivity and performance at the level of individuals, but rather to distribute funding between institutions. And even in this context, it constitutes but one small part of a larger funding mechanism. Indicators of this type are seldom decisive to the overall budget distributions to higher education institutions (Zacharewicz et al. 2019). In Norway, only about 2 per cent of the total funding for the university sector is distributed based on NPI results, making its distributional effect both modest and predictable (Aagaard 2018). Thus, it would seem to make no real difference at all.

However, when used at the individual level, its effect may be anything but modest and predictable. There are several examples in Europe of how similar indicators are used to evaluate and reward individual researchers (Aagaard 2015; Cleere and Ma 2018; Hammarfelt 2018; Mingers and Leydesdorff 2015). The seriousness and scope of this kind of problem is reflected in the Declaration on Research Assessment (DORA) and the Leiden Manifesto, which specifically recommend against using journal-based metrics to assess individual researchers (Hicks et al. 2015). From a gender perspective it is such local use, where individuals are evaluated based on their production of 'publication points' rather than their research, that the indicator represents a problem because of the way it disproportionately rewards men and reifies gendered notions of excellence. These gendered notions of excellence are commonly carried over into biases affecting hiring or promotion contexts (Coate and Howson 2016; O'Connor and O'Hagan 2015; van den Brink and Benschop 2012).

Many other countries have adopted an indicator like the NPI, either at country level or at single universities, but mostly with local modifications (Zacharewicz et al. 2019). For

example, some Swedish universities give credit to local book series (Hammarfelt 2018); and the University College in Dublin, in addition to weighting book chapters and journal articles equally, also includes edited books and published reports (Cleere and Ma 2018). Our findings suggest that greater scholarly attention should be paid to the way these different modifications might affect gender gaps in productivity.

The sensitivity of the gender gap to the ways in which productivity is measured also suggests that adjustments may be made to the NPI to help neutralize these amplifying effects. For example, the NPI uses a multiplicative method that takes a point of departure in each publication separately where the value of each component depends on the value of the others. The points awarded for each publication are first fractionalized by author contributions, then weighted with respect to publication type, then publication channel, and finally by international collaboration. In contrast, one could imagine a system of non-multiplied components, where each component has a pre-determined share of the overall score. Modifying the types of publications included (as well as the way they are weighted) might also have a mitigating effect. However, as pointed out by Wilsdon et al. (2016), 'One size is unlikely to fit all: a mature research system needs a variable geometry of expert judgement, quantitative and qualitative indicators.' Perhaps the best way forward is to ensure that the NPI is used in combination with other forms of assessment.

In conclusion, this study suggests that the choices made in constructing an indicator can affect how gender differences in productivity are perceived. Our study shows that even setting aside gendered expectations for performance (which are difficult to capture in a quantitative study), the skewed distribution of women in the academy coupled with field-related diversity in publication practices means that indicators benefiting male-dominated fields may inadvertently portray women as less productive than men. The way the NPI measures productivity has led to an identifiable gender-based Matthew effect, showing that although the concept of 'excellence' may be gender neutral in theory, it appears to benefit men in practice.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Notes on contributors

Lynn P. Nygaard is a special advisor at the Peace Research Institute Oslo (PRIO), Norway, where she helps researchers publish academically, secure grants and develop as professionals. She holds a BA from the University of California, Berkeley, in women's studies; a graduate degree in political science from the University of Oslo; and a Doctor in Education (EdD) from the Institute of Education, University College London. As a researcher, she uses ethnographic methods and bibliometric analyses to better understand the different contexts of academic writing and the wide variety of publishing practices, including sites of negotiation women face on the path to professorship.

Fredrik Niclas Piro, PhD, is a research professor at the Nordic Institute for Studies in Innovation, Research and Education (NIFU). His research interests cover bibliometrics (including gender analyses), research funding analyses and R&D indicator development, particularly related to health and medical research. He also participates in an on-going project on the lack of gender balance in academia (BALANSE).

Dag W. Aksnes is research professor at the Nordic Institute for studies in Innovation, Research and Education (NIFU) and affiliated with the Centre for Research Quality and Policy Impact Studies (R-QUEST). He holds a PhD from the University of Twente, the Netherlands. Aksnes' research covers various topics within the field of bibliometrics, including studies of gender differences. He coordinates an ongoing research project on the lack of gender balance in academia (BALANSE).

References

Aagaard, K. 2015. "How Incentives Trickle Down: Local use of a National Bibliometric Indicator System." *Science and Public Policy* 42 (5): 725–737.

Aagaard, K. 2018. "Performance-based Research Funding in Denmark: The Adoption and Translation of the Norwegian Model." *Journal of Data and Information Science* 3 (4): 20–30.

Aagaard, K., C. Bloch, and J. W. Schneider. 2015. "Impacts of Performance-Based Research Funding Systems: The Case of the Norwegian Publication Indicator." *Research Evaluation* 24: 106–117.

Aksnes, D. W., and G. Sivertsen. 2019. "A Criteria-Based Assessment of the Coverage of Scopus and Web of Science." *Journal of Data and Information Science* 4 (1): 1–21.

Baker, M. 2010. "Choices or Constraints? Family Responsibilities, Gender and Academic Career." *Journal of Comparative Family Studies* 41 (1): 1–18.

Cameron, E.Z., M.E. Gray, and A. M. White. 2013. "Is Publication Rate an Equal Opportunity Metric?" *Trends in Ecology & Evolution* 28(1): 7-8.

Clarence, S., and S. McKenna. 2017. "Developing Academic Literacies Through Understanding the Nature of Disciplinary Knowledge." *London Review of Education* 15 (1): 38–49.

Cleere, L., and L. Ma. 2018. "A Local Adaptation in an Output-Based Research Support Scheme (OBRSS) at University College Dublin." *Journal of Data and Information Science* 3 (4): 74–84.

Coate, K., and C. K. Howson. 2016. "Indicators of Esteem: Gender and Prestige in Academic Work." *British Journal of Sociology of Education* 37 (4): 567–585.

De Bellis, N. 2014. "History and Evolution of (Biblio)Metrics." In *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, edited by B. Cronin, and C. R. Sugimoto, 23–44. Cambridge: The MIT Press.

Engels, T. C. E., and R. Guns. 2018. "The Flemish Performance-Based Research Funding System: A Unique Variant of the Norwegian Model." *Journal of Data and Information Science* 3 (4): 45–60. doi:10.2478/jdis-2018-0020.

Furner, J. 2014. "'The Ethics of Evaluative Bibliometrics'." In *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, edited by B. Cronin, and C. R. Sugimoto, 85–107. Cambridge: The MIT Press.

Gingras, Y. 2014. "Criteria for Evaluating Indicators." In *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, edited by B. Cronin, and C. R. Sugimoto, 109–125. Cambridge: The MIT Press.

Hammarfelt, B. 2018. "Taking Comfort in Points: The Appeal of the Norwegian Model in Sweden." *Journal of Data and Information Science* 3 (4): 85–95.

Hicks, D., P. Wouters, L. Waltman, S. de Rijcke, and I. Rafols. 2015. "Bibliometrics: The Leiden Manifesto for Research Metrics." *Nature* 520 (7548): 429–431.

Kyvik, S. 1991. *Productiviy in Academia. Scientific Publishing at Norwegian Universities*. Oslo: Universitetsforlaget.

Lea, M. R., and B. V. Street. 2006. "The 'Academic Literacies' Model: Theory and Applications." *Theory Into Practice* 45 (4): 368–377.

Lillis, T., and M. Scott. 2007. "Defining Academic Literacies Research: Issues of Epistemology, Ideology and Strategy." *Journal of Applied Linguistics* 4 (1): 5–32.

Lipton, B. 2015. "A new 'ERA' of Women and Leadership. The Gendered Impact of Quality Assurance in Australian Higher Education." *Australian Universities' Review* 57 (2): 60-70.

Lund, R. 2020. "The Social Organisation of Boasting in the Neoliberal University." *Gender and Education* 32 (4): 466–485.

Mingers, J., and L. Leydesdorff. 2015. "A Review of Theory and Practice in Scientometrics." *European Journal of Operational Research* 246 (1): 1–19.

Moore, S., C. Neylon, M. P. Eve, D. P. O'Donnell, and D. Pattinson. 2017. "'Excellence R Us': University Research and the Fetishisation of Excellence." *Palgrave Communications* 3 (16105): 1–13.

Moreley, L. 2016. "Troubling Intra-Actions: Gender, neo-Liberalism and Research in the Global Academy." *Journal of Education Policy* 31 (1): 28–45.

Nygaard, L. P. 2017. "Publishing and Perishing: An Academic Literacies Framework for Investigating Research Productivity." *Studies in Higher Education* 42 (3): 519–532.

Nygaard, L. P., and R. Bellanova. 2018. "'Lost in Quantification: Scholars and the Politics of Bibliometrics.'." In *Global Academic Publishing: Policies, Perspectives and Pedagogies*, edited by M. J. Curry, and T. Lillis, 23–36. Bristol: Multilingual Matters.

O'Connor, P., and C. O'Hagan. 2015. "Excellence in university academic staff evaluation: a problematic reality?" *Studies in Higher Education* 41(11): 1943-1957.

Hug, S. E., M. Ochsner, and H.-D. Daniel. 2014. "A Framework to Explore and Develop Criteria for Assessing Research Quality in the Humanities." *International Journal for Education Law and Policy 10*(1): 55-64.

Piro, F.N., Aksnes, D.W. and Rørstad, K. (2013): "A Macro Analysis of Productivity Differences Across Fields: Challenges in the Measurement of Scientific Publishing." *Journal of the American Society for Information Science and Technology* 64(2):307-320.

Pölönen, J. 2018. "Applications of, and Experiences with, the Norwegian Model in Finland." *Journal of Data and Information Science* 3 (4): 31–44. doi:10.2478/jdis-2018-0019.

Rørstad, K., and D. W. Aksnes. 2015. "Publication Rate Expressed by age, Gender and Academic Position - A Large-Scale Analysis of Norwegian Academic Staff." *Journal of Informetrics* 9: 317–333.

Sivertsen, G. 2018. "The Norwegian Model in Norway." *Journal of Data and Information Science* 3 (4): 3–19.

Smith, D. 2017. "Progress and Paradox for Women in US Higher Education." *Studies in Higher Education* 42 (4): 812–822.

Squazzoni, F., G. Bravo, F. Grimaldo, D. García-Costa, M. Farjam, and B. Mehmani. 2021. "Gender gap in Journal Submissions and Peer Review During the First Wave of the COVID-19 Pandemic. A Study on 2329 Elsevier Journals." *PLoS ONE* 16 (10): e0257919.

Vabø, A., H. Gunnes, C. Tømte, A.C. Bergene, and C. Egeland. 2012. *Kvinner og Menns Karriereløp i Norsk Forskning: En Tilstands Rapport [Women and Men's Career Trajectories in Norwegian Research: A Status Report]*, Vol. 2019-9, 85. https://nifu.brage.unit.no/nifu-xmlui/handle/11250/280862

van den Brink, M., and Y. Benschop. 2012. Slaying the Seven-Headed Dragon: The Quest for Gender Change in Academia, *Gender, Work & Organization* 19(1): 71-92.

Wilsdon, J., et al. 2015. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, 163. doi: 10.13140/RG.2.1.4929.1363

Witt, C. 2011. "'What is Gender Essentialism?'." In *Feminist Metaphysics. Explorations in the Ontology of Sex, Gender and the Self. Feminist Philosophy Collection*, edited by C. Witt, 11–25. Dordrecht: Springer.

Zacharewicz, T., B. Lepori, E. Reale, and K. Jonkers. 2019. "Performance-based Research Funding in EU Member States—a Comparative Assessment." *Science and Public Policy* 46 (1): 105–115.