4-16-2019

Evaluation complacency or evaluation inertia? A study of evaluative metrics and research practices in Irish universities

Lai Ma
University College Dublin, lai.ma@ucd.ie

Michael Ladisch
University of the Pacific, mladisch@pacific.edu

Recommended Citation
Ma, L., & Ladisch, M. (2019). Evaluation complacency or evaluation inertia? A study of evaluative metrics and research practices in Irish universities. *Research Evaluation, Early online publication*, DOI: 10.1093/reseval/rvz008
https://scholarlycommons.pacific.edu/libraries-articles/111

Evaluation Complacency or Evaluation Inertia? A Study of Evaluative Metrics and Research Practices in Irish Universities

Lai Ma*
University College Dublin, Belfield, Dublin, Ireland
Email: lai.ma@ucd.ie

Michael Ladisch
University of the Pacific Stockton, California, USA
Email: mladisch@pacific.edu

*Corresponding author

Acknowledgements
The authors would like to thank the reviewers for their invaluable comments and suggestions. They also thank the participants of this study for their time, insights, and support.

Evaluation Complacency or Evaluation Inertia? A Study of Evaluative Metrics and Research Practices in Irish Universities

Abstract

Evaluative metrics have been used for research assessment in most universities and funding agencies with the assumption that more publications and higher citation counts imply increased productivity and better quality of research. This study investigates the understanding and perception of metrics, as well as the influences and implications of the use of evaluative metrics on research practices, including choice of research topics and publication channels, citation behaviour, and scholarly communication in Irish universities. Semi-structured, in-depth interviews were conducted with scholars and researchers from the humanities, the social sciences, and the sciences in various career stages. Our findings show that there are conflicting attitudes towards evaluative metrics in principle and in practice. The phenomenon is explained by two concepts: evaluation complacency and evaluation inertia. We conclude that evaluative metrics should not be standardized and institutionalized without a thorough examination of their validity and reliability and without having their influences on academic life, research practices, and knowledge production investigated. We also suggest that an open and public discourse should be supported for the discussion of evaluative metrics in the academic community.

Keywords

Evaluative Metrics, Research Assessment, Research Practices, Citation Behaviour, Evaluation Complacency, Evaluation Inertia

1. Introduction

Increasingly, indices such as the journal impact factor, h-index, and citation counts are used by universities and funding agencies as "objective measures" of research quality and output. Such evaluative measures are integral elements in the decision-making process of appointment and promotion of academic positions as well as grant applications. Nevertheless, their validity and reliability have not been agreed upon in the academic literature (see, for example, Arnold & Fowler, 2011; Weingart, 2005). Recent publications such as the San Francisco Declaration for Research Assessment, commonly known as DORA (ASCB, 2013), The Metric Tide (Wilsdon et al., 2015), and the Leiden Manifesto for Research Metrics (Hicks et al., 2015) caution against the blind use of evaluative metrics in research assessment. These calls for responsible metrics argue that evaluative metrics do not sufficiently represent the quality of research and that they are not comparable across disciplines and research fields.

The rationale for the use of evaluative metrics in research assessment is, presumably, meritocracy: the higher the citation counts, the higher the quality of research. Their use for research assessment has become commonplace with the development and provision of various measures including the journal impact factor, h-index and citation counts by proprietary providers, as well as various frameworks in performance-based funding systems (Hicks, 2012; Zacharewic, Lepori, Reale & Jonkers, 2018). These measures are often used by universities for recruitment and promotion and by funding agencies for evaluating grant applications. Moreover, evaluative metrics are also used for ranking universities worldwide. For example, citations per faculty account for 20% of the overall ranking score in the QS

Universities Ranking (QS World University Rankings – Methodology, n.d.). Notably, the criteria for evaluation of university performance are almost exclusively bibliometric in the Shanghai Ranking, prepared annually by the Shanghai Jiao Tong University: the calculation is based on the total number of Nobel Prizes and Fields Medals for researchers and alumni (30%), number of highly cited researchers (20%), papers published in the journals Nature and Science (20%), papers published in SCI and SSCI indexed journals (20%), and institution's per capita academic performance with regard to all indicators listed (10%) (Ranking Methodology of Academic Ranking of World Universities – 2016, n.d.).

Furthermore, some performance-based funding systems such as the Norwegian model use metrics for allocating block grants to universities. Although this type of metrics is not designed to evaluate individuals, there is evidence that the incentives might trickle down and affect research practices (Aagaard, 2015).

To what extent are evaluative metrics actually used in contexts such as promotion and funding applications, however, is relatively unknown except for some performance-based funding systems. That is to say, metrics are usually used with a basket of criteria and how much each component is considered or counted is not clear. Nevertheless, the requirement of including metrics such as citation counts and h-index in applications certainly affects the perception of their significance in research assessment, which would in turn influence how academics work, including how academics select research topics, publication channels, collaborators, and so on.

In Ireland, evaluative metrics are used as a criterion in research evaluation in universities and funding agencies, although the practices are not standardised. Irish universities go through quality review regularly under the Qualifications and Quality Assurance (Education and Training) Act 2012. These reviews often involve presenting profiles of staff and their research outputs, including metrics obtained from various providers, to internal and external examiners. It is uncertain, however, as to how important metrics are in the review process. The question goes the same in funding applications, in which metrics including the number of publications and citation are required, yet how much they weigh in an application is not clearly stated.

It is also commonly expected that evaluative metrics would be considered in the processes of recruitment and promotion of academic staff in all disciplines. From informal conversations with Irish academics, criteria for research evaluation is often determined in situ, not unlike the customary rules in Lamont's (2009) study of panels in grant agencies. At the time of this writing, there is not a national research evaluation framework in Ireland. Output-based research support scheme has been implemented in one of the Irish universities for two years (Cleere & Ma, 2018), and two other universities are considering rolling out similar systems in the coming year. However, the schemes are not intended to use as tools for evaluation, but rather, incentives for increasing research productivity and quality. Notwithstanding research evaluation is not standardised in Ireland, the importance of metrics such as journal impact factor and citation counts have been promoted through venues such as library workshops, for example, "how to track your citations" and "how to increase visibility of your research". Some research offices also promote the benefits of publishing in journals of high impact factor. There is certainly an effort to raise awareness of evaluative metrics as a representation of research performance. The perception and understanding of evaluative metrics among Irish academics, however, have not been studied.

Using in-depth semi-structured interviews, this study investigates the understanding and perceived importance of evaluative metrics by academics in Ireland, as well as the implications on research practices, including selection of research topics and publication channels, citation behaviour and scholarly communication, where a national or institutional evaluation systems are non-existent and the importance and use of evaluative metrics are not explicitly stated and standardised. The study shows that there are conflicting attitudes between the participants' view of evaluation metrics in principle and in practice. In principle, they voiced the limitations of evaluative metrics as a tool for research assessment. In practice, they actively use metrics as a guide to determine productivity of individuals and quality of research outputs. The conundrum is manifested in their publication and citing practices, as well as their opinions as to whether evaluative metrics should be used for research assessment. In the discussion section, we will explain the discrepancy using the concepts of evaluation complacency and evaluation inertia, indicating the need for 'metrics literacy' among academics, as well as an open and public discourse about evaluative metrics.

2. Literature Review

Evaluative metrics are commonly understood as the quantitative measures of research outputs and publications, primarily based on citation-based data. While the original idea of the citation index was to provide subject control of the literature (Garfield, 1955), it has been used for evaluating the impact of journals (Garfield, 1972) and various kinds of analyses based on citation data, from co-citation analysis (e.g., White & McCain, 1998), to interdisciplinarity (e.g., Porter & Rafols, 2009), to the development and growth of disciplines (e.g., Lariviere, Sugimoto, & Cronin, 2012).

However, the increasing use of evaluative metrics in research assessment has raised concerns since citation analysis is not a method without flaws and limitations due to factors such as negative citations and self-citations (see, for example, MacRoberts & MacRoberts, 1989), not to mention that the measures do not take into account research practices and epistemic cultures in different disciplines (Bornmann & Leydesdorff, 2014). Researchers in various disciplines have questioned the use of bibliometrics as evaluative criteria of research performance (for example, Blockmans, Engwall & Weaire, 2014; Blyth, et al., 2010; Erne, 2007; Lawrence, 2007; Van Dalen & Henkens, 2012).

The use of evaluative metrics for research assessment has been critiqued and questioned in other contexts including the discussion of the idea and ideals of universities (see, for example, Readings, 1996; Clark, 2006; Collini, 2012), specifically, the use of these metrics implicated in managerialism and commercialisation of universities (see, for example, Parker, 2011; Parker & Jary, 1995; Clarke & Knights, 2015), audit cultures (see, for example, Amit, 2000; Craig, Amernic & Tourish, 2014; Hammarfelt & de Rijcke, 2015; Shore, 2008), and the culture of speed (see, for example, Berg & Seeber, 2016; Vostal, 2015). Most importantly, it has been noted that a system of evaluation could weaken critical voices in academia. Sosteric (1999), in particular, has stated that "as time passes, and graduate students do what they have to get hired on, less and less of the essence of critical thinking and critical discourse will be rediscovered each generation. It will be a gradual devastation, but it will be devastation nonetheless." Cronin (2005) has also noted the extensive use of evaluative metrics may lead to "symbolic capitalism." In the edited volume *Beyond Bibliometrics*, Day (2014) questions the commodification of human relationships and communication in the use of evaluative metrics, on the one hand, and Furner (2014) calls for ethical use of evaluative metrics, on the other. Publications such as The Leiden Manifesto (Hicks, et al., 2015) and The Metric Tide (Wilsdon, et al., 2015) call for responsible use of metrics for evaluating research outputs.

In the past few decades, there have been contributions stating the importance of studying cultures and politics in citation analysis and related topics (see, for example, Cronin, 2005; Weingart, 2005; Wouters, 1999). Conceptually, Dahler-Larsen (2012) suggests using the concept of 'evaluation machine' to understand how people have internalised evaluation as routines and themselves as object of evaluation. Recent studies about the effects of the use of evaluative metrics on research practices have been systematically reviewed by de Rijcke et al. (2016). Butler (2003, 2004), for example, has found that academics respond to metrics with strategic behaviour, specifically, increased publications with lower impact factor while striving less for high quality journals that require more investment as a result of performance-based funding system implemented in Australia, where raw number of publications, regardless of impact factor, quality, and prestige, were counted. Aksnes & Rip (2009) investigate how Norwegian scientists perceive citations in terms of quality and visibility, as well as their legitimacy and conclude that scientists' perceptions are 'ambivalent' in the sense that they are irresolute about the validity and legitimacy of citations in research practices and assessments. Most recently, Müller & de Rijcke (2017) report that researchers in the life sciences have been "thinking with indicators" in their research activities and that metrics-driven research has been routinised in knowledge production. However, citing behaviour has largely been studied pertaining to the types and motivation of citation (Bornmann & Daniel, 2008).

Few have, however, studied how the understanding and perceived importance of evaluative metrics related to everyday research practices, including the choice of research topics and scholarly communication. Wittgenstein (1958), in exploring language use, states: "And to *think* one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately': otherwise thinking one was obeying a rule would be the same as obeying it" (§202, emphasis in original). The production and use of evaluative metrics are based on the norms and rules established through day-to-day interactions of funding agencies, university management, academics and other stakeholders. The ongoing processes of the production and use of evaluative metrics hence shape and reshape the norms and rules in epistemic cultures. The objective of this study is to understand the implications and consequences of evaluative metrics with regard to research practices, including citation behavior and scholarly communication, particularly in the context of Ireland, where the importance of metrics is not explicit and standardized in national or even institutional levels. Specifically, what is the perceived importance of evaluative metrics such as impact factor and h-index? What are the changes, if any, of research practices and citation behaviour with the increasing use of evaluative metrics among academics in Ireland?

3. Method

The study was conducted in Ireland between May 2016 and June 2017. In-depth, semi-structured interviews of 40-70 minutes were conducted with a total of nineteen participants. A pilot study was conducted with four participants. Thereafter the interview questions were revised to elicit more relevant responses pertaining to the research questions. After the pilot study, invitations for interview were sent in two phases. In the first phase, invitations were sent to all school/department managers in all research universities by email. Unfortunately, only one response was received from a doctoral student, hence no interview was conducted. In the second phase, an email list was created by collecting individual email addresses using staff directories on university webpages, including all who were listed as academic or research staff. The potential participants were then invited to participate by email. To balance the number between different disciplines, repeated prompts were sent to potential

participants in the humanities in the last phase of the study. In the interviews, the participants were asked to identify their research in the sciences, the social sciences, the humanities, or other. Seven of the participants identified themselves in the sciences, seven in the social sciences, four in the humanities, and one participant identified herself in the overlapping area of the social sciences and the humanities. The participants hold academic or research positions from post-doctoral researchers to full professors.

3.1 Semi-Structured Interviews and Data Analysis

The interviews covered three major areas: (a) the participant's understanding and use of evaluative metrics, (b) the participant's publication and citation practices, and (c) the participant's opinions on the usefulness and appropriateness of the use of evaluative metrics in research assessment.

In the first part of the interview, the participants were asked to describe their understanding and use of evaluative metrics. The question was used to gauge the participants' knowledge and perception about the measures, providers, and uses of evaluative metrics. In the second part, the participants were asked to describe their publication and citation practices based on their perceived importance of evaluative metrics in research assessment, including topics such as choice of publication channels and research topics, dissemination of research output, and so on. In the last part of the interview, the participants were asked to comment as to how evaluative metrics should or should not be used in research assessment.

All interviews were audio-recorded, de-identified and transcribed for data analysis. The transcripts were analysed separately by the two authors of this article. The development of topical categories was a dynamic process using topics and key terms emerged from the interview transcripts. The categories of topic were then compared and discussed. The findings were consistent between the two data analyses.

4. Findings

Evaluative metrics are understood by the participants as quantitative measures to represent the quality of research output. Journal impact factor, citation counts, and h-index were mentioned. The ResearchGate Score (hereafter "RG Score") was also recognised as evaluative metrics. However, none seemed to be familiar with metrics such as SJR, SNIP or Eigenfactor; altmetrics were largely unknown. It is also clear that the participants were not fully informed about what these metrics represent. There are also conflicting attitudes toward evaluative metrics in principle and in practice: The importance of evaluative metrics perceived by the participants depicts a complex picture of not only the use of evaluative metrics for research assessment, but also the cultural milieu of academic and research institutions.

On the one hand, the participants recognised that metrics are increasingly used for ranking applications of academic appointments, promotion, and grant applications before the peer review process and they were in agreement that evaluative metrics should not be used as the sole criterion for assessing research performance. There is a sense of resentment of how evaluative metrics are increasingly used as administrative and managerial tool. Their impression is that many administrators or funding agencies do not understand the nature of metrics but adopt them conveniently in the name of fairness, objectivity and accountability. The participants noted the use of evaluative metrics as part

of "how the system works" and hence they must "play the game." It is notable that half of the participants used the terms "game" or "gaming" during the interviews.

On the other hand, the participants keep track of their own metrics including citation counts, h-index and RG Score and stated that metrics are a motivating factor in their work. For the participants in senior positions, they gain a sense of achievement when their metrics go up. For the participants in their early-career positions, the metrics are much less of an affirmation but a goal to achieve. They most often compare their metrics with colleagues when applying for academic positions or promotion, while some put it simply as a matter of ego. Their own uses of evaluative metrics implicate a level of acceptance of citation counts, h-index and RG Scores as standards for assessing and comparing the quality of research outputs.

In principle, the participants distrust evaluative metrics as objective measures of research performance, claiming that quantitative measures do not and cannot replace expert judgement and peer review. In practice, they are active users of these metrics for personal and/or administrative purposes. Nevertheless, using metrics for research assessment is perceived as a standardised practice in academia, notwithstanding the participants' opinions about the validity, reliability and legitimacy of the various measures. The perception of evaluative metrics use as a standardised practice is also manifested in their discussion of publication and citation practices.

4.1 Publication Practices

> "It certainly has impact… I have done some educational research, not published, because the impact factor is so low, that you can't… the effort to do a paper in physics and the effort to do a paper in educational research. The impact factor is hugely different." (Participant #2 – Sciences – Associate Professor)

> "The paper that I can finish the quickest and would have the biggest impact becomes more important than the paper I am passionate about… The system rewards short-termism." (Participant #10 – Social Sciences – Associate Professor)

Prompted by the question about the relationship between evaluative metrics and publication practices, the participants respond with a range of topics in which evaluative metrics are considered, from epistemic cultures, collaboration, supervision and mentoring, to research funding and university management. The participants in the Sciences are more conscious of evaluative metrics in their publication practices and are well-versed in journal impact factor. They are also more attentive to metrics such as citation counts and h-index since the metrics are often required in grant applications. While they emphasised that their choices of research topic and research project are primarily driven by their research interests, the availability of funding opportunities is a determining factor in their research and publication plans. Since it is not clear how important the track record, such as the number of publications, citation counts, and h-index, are in grant applications, there is a perception that the higher the metrics, the better chance of winning a grant, which in turn leads to some changes in publication practices, including the choice of 'hot topics' and the so-called "salami-slicing model". For instance, a participant noted that the number of publications is crucial for increasing the h-index,

strategies are hence adopted: "*The fact is, you know, if you publish two small works, you get more citations than if you publish a comprehensive work*." (Participant #4 – Sciences – Assistant Professor).

Publication practices are also affected by the participants' role as supervisors and mentors. Considering that academic positions are highly competitive, and that the number of publications and citation counts in some disciplines would be the first criteria for determining the eligibility of candidacy, the participants in senior positions consider co-authorship as a duty to help doctoral students and post-doctoral colleagues working in their projects: "*You know you keep on especially when you have a PhD student coming to an end. I try that each PhD student in the group has 2 or 3 publications at the time when the thesis is submitted. This is the priority. When there's some time left, I can continue on the book, in the last year, it never happened*." (Participant #3 – Sciences – Professor).

The participants in the Social Sciences and the Humanities focus more on the prestige and reputation of journals and publishers recognised by peers in their disciplines and placed less importance on evaluative metrics than their counterparts in the sciences. They are increasingly attentive to journal impact factor and citation counts due to the availability of data on platforms such as Google Scholar and ResearchGate. Reporting requirements by university management and quality assurance exercise have also raised the awareness of evaluative metrics. The participants noted that the high expectation of number of publication in recruitment and promotion has led some to publish at a faster rate. Research topics and projects that take longer time could be abandoned as a result. However, the influences of evaluative metrics such as citation counts and h-index on publication practices are not as prevalent as in the Sciences.

4.2 Citing Practices

> "If you are interested in gender equality people have noticed that men are more often cited than women, so now if you want to try to help other women, you try to think of people in the profession who have written relevant articles…. These are the people that you could cite to increase the visibility of women to counteract that imbalance. So that means you might be more inclined to cite more women now to improve their rankings." (Participant #16 – Humanities/Social Sciences – Associate Professor)

> "In the Humanities we often cite work that's quite out of date, as part of a discursive rhetorical device to develop an argument. We might even cite things we extensively disagree with. Just a way to have an opposition and a dialectical process." (Participant #15 – Humanities – Assistant Professor)

Diverse publication practices have been widely discussed in terms of publication types and frequency, citing practices, however, have not been thoroughly examined. The participants in the Humanities in this study shed light on the 'impossibility' of using citation-based metrics for research evaluation in their disciplines, not only because some of them publish in monographs compared to journal articles or conference proceedings, but also their citing practices are very different from research studies in

the sciences and the social sciences. The participants mentioned the following reasons: first, works in the Humanities are not like building blocks, that is, one's work doesn't necessarily build on one another. A dialogue can be made between a 13th century historical or philosophical piece and no other references would be made. Second, lengthy, in-depth book reviews are still seen as the most important 'indicator' of the quality of work in some disciplines. The number of reviews, instead of citations, signifies the attention by and engagement with the scholarly community. Third, influential works are recognised as they become canonical, meaning that they have the status of a "required reading" in the field, albeit they are not necessarily cited. Therefore, citation-based metrics are not considered appropriate for determining the quality or significance of scholarly work in the Humanities.

Self-citations and exchange of citations are described as gaming behaviour by the participants in the Social Sciences and the Sciences. The practices of self-citations or 'citation requests' are perceived as a consequence of citation-based metrics. Interestingly, gaming behaviour is adopted by one participant as a way to increase the ranking and visibility of female scholars. The participants also recited instances being asked to add references in journal articles under review. Notwithstanding the motivation of gaming behaviour, it is based on the assumption that citations are considered in research assessment. Citing practices also adapt to the larger system of academic cultures in which evaluative metrics are perceived to play an important role in research assessment.

Acknowledging that citation data do not necessarily represent the quality of research outputs, and that citations can be gamed, the participants in the Social Sciences and the Sciences noted that, however, they sometimes include highly cited articles in the background section of their writings, partly because Google Scholar ranks search results based on citation counts. When doing so, they trust that a highly-cited paper can be treated as a good source, although they are more critical when citing references about the specific topic of their research. As one participant noted: "*I have to find this reference. Who did we use last time, who somebody else uses? Some of them you haven't, honestly you haven't seen the original. The results go back to… in 1906. That paper in the library somewhere and I am not going there. So I am just going to trust the other 40 people who have already cited this paper. That's the… It's lazy and it's probably not the academic way of doing things. But time is short so you would just trust that this is the important one because everyone else has cited it.*" (Participant #6 – Sciences – Assistant Professor).

On the one hand, citing practices, in addition to publication types and frequency, explain the inappropriateness of the use of citation-based evaluative metrics in the Humanities. Citations are not meaningful in terms of quality or impact in their research practices. Nevertheless, the use of evaluative metrics by universities and funding agencies is reshaping publication and citation practices. The implications on knowledge production and epistemic cultures would need examination over time.

On the other hand, citing practices are questioned by participants in the Social Sciences and the Sciences based on their own experiences of being asked to include specific references by reviewers, as well as famous cases of misuse of citations for increasing journal impact factor (see, for example, Davis, 2008). Yet, they trust that the number of citations represents the quality of a journal article, especially when they need background information for their writings. In sum, they rely on citations as objective measures when they are looking for sources, although they are critical of citation-based metrics because of disciplinary differences and gaming behaviour observed.

4.3 Evaluative Metrics for Research Assessment

> "I do mostly pure maths, I also do some of the applied things. I have one small part in a paper that is in the engineering journal. The engineering journal has five times of impact factor as any of the mathematical journals. Therefore, that's my best paper, by having nothing, no real insight of mine in it. If I were asked what I do, I wouldn't mention that at all." (Participant #6 – Sciences – Assistant Professor)

> "I tend to be very mistrustful of citation index and things like that. They don't serve us very well. If you are an active researcher, you know why some work is cited, I don't need data to tell me." (Participant #17 – Humanities – Professor)

The participants' criticisms of evaluative metrics as a research assessment tool and their own active use should be considered with their opinions as to whether evaluative metrics should be used in academia. Two participants from the Humanities and one in the Sciences voiced total objections of the use of evaluative metrics. For them, the use of evaluative metrics is not justified and is deemed totally inappropriate in their research areas as they regard peer review as the most important mechanism in assessing the quality of scholarly work.

Most participants, however, are not against the use of evaluative metrics, and suggested that metrics can be used in objective manner insofar as disciplinary differences and epistemic cultures are taken into account. Some also suggested that evaluative metrics could prevent dogmatism and conservatism in the peer review process. The resistance of the use of evaluative metrics is primarily based on their views that current metrics in use should be fine-tuned and improved before they are standardised and used for hiring, promotion, and grant applications across disciplines. Furthermore, the participants were also concerned about the effects of over-emphasising evaluative metrics in academia, including gaming behaviour, irresponsible management and leadership, and academic integrity.

Another 'side-effect' of using evaluative metrics for research assessment noted by the participants is the proliferation of publications, which also affects the quality and speed of peer review. It is perceived that the pressure to publish has led to increased number of submissions. At the same time, academics can spare little time to serve as reviewers, particularly when reviewing is not recognised and valued in promotion and grant application processes.

5. Discussion

This study shows that there are conflicting attitudes towards evaluative metrics in principle and in practice. Although the participants are sceptical about the representativeness and objectivity of evaluative metrics, they use metrics to make judgements about quality and trustworthiness of articles, as well as productivity of colleagues. The perception of evaluative metrics and their importance, however, is not ambivalent as in Aksnes and Rip's (2009) study. The participants provided reasons behind their resistance and objection to the uses of evaluative metrics by the "system"—the decision-making processes in hiring, promotion, and funding applications—including the inappropriateness of using metrics for comparing and ranking across disciplines and research specialisations. The participants also provided suggestions with regard to the development and improvement of evaluative

metrics and the degree to which evaluative metrics should be used in research assessment, albeit they are less critical of their own uses of evaluative metrics for academic writing and their tracking of own and others' citation-based metrics. Aiming for higher journal impact factor or citation counts has become a form of academic life to the extent that its objectivity is presumed. "Thinking with indicators" has become routine practices of knowledge production (Müller & de Rijcke, 2017). The findings also reflect the discussion of the "reification of evaluative indicators" in that the importance of evaluative metrics is deep-seated and have been firmly established (de Rijcke, et. al, 2016).

In other words, evaluative metrics are perceived as rules and standards in everyday academic life. They are used 'subconsciously' in research activities, as if academics have accepted the legitimacy of metrics as measures of research quality. When they were asked to discuss the use of metrics, however, the participants were reflective about many aspects of metrics. Why is their thinking about metrics in principle and their action in practice inconsistent? Is the acceptance of the use of evaluative metrics due to the fear of sanction and repercussion as some studies in new public management suggest? Or is it because there is a lack of open discourse in academia about evaluative metrics for generating norms and rules of research assessment?

We should take note of two issues in this conundrum. Firstly, there seems to be a lack of understanding of evaluative metrics among academics in Ireland. Although the participants mentioned journal impact factor, h-index, and RG scores, the majority cannot tell how these metrics are calculated or what they represent. The participants' dissatisfaction of evaluative metrics was triggered when they discussed metrics being used as administrative or managerial tools, but most did not discuss their concerns about the objectivity or validity of these metrics. Rather, they elaborated normative issues such as appropriateness and fairness with little knowledge about how these indices are calculated. This lack of basic understanding can lead to folk theories that are not fruitful for developing guidelines for responsible use. Academics and other stakeholders should be educated and informed as to why certain metrics is appropriate or inappropriate in different contexts, as 'metrics literacy' would be a necessary requirement for using metrics responsibly. For example, if doctoral students are educated about evaluative metrics—what they are, what they actually measure—then they will make more informed decision in their research activities rather than blindly accept the 'rules' or folk theories about metrics when they move through the academic lifeworld and, more importantly, when they participate in decision-making processes of various academic activities, including recruitment and promotion and strategic planning.

Secondly, the conundrum implicates the lack of open discourse about evaluative metrics within the academic environment in Ireland. Public forums in which academics can discuss ideas and opinions about general academic issues such as evaluative metrics are not commonplace. The participants did not assume an active role in the discourse about metrics even though they were not satisfied with the fairness and appropriateness. We explain the lack of discourse using two concepts: *evaluation complacency* and *evaluation inertia*.

- Evaluation complacency: when one is complacent about the achievements measured by evaluative metrics, not feeling the need to reflect on the limitations and shortcomings of metrics. Implicit in evaluation complacency is the acceptance of metrics to be used as standards. For example, if an academic has been rewarded by playing the rule of the games (regardless of its value or legitimacy), he or she would not want to change the status quo. Rather, he or she would reinforce the use of metrics. The danger of complacency is a system where the rich gets richer and hence reinforces the system without reflection and critique.

- Evaluation inertia: when there is no tendency to reflect, critique, and change the existing standards of research evaluation, including the use of metrics. There are different reasons and circumstances that could lead to evaluation inertia. For example, a system imposes severe sanction upon certain behaviour and action, signaling things cannot be changed. Evaluation inertia can also be due to the chasing of metrics being incentivised. As such, when competition intensifies (e.g., one needs a high number of publications to secure an academic position), there is no time or head space to reflect and critique existing standards and practices, which in turn leads to the lack of discourse about metrics, among other things.

Evaluation complacency does not explain the conflicting attitudes of the participants in this study, as they were aware of the limitations of metrics and the necessity to deal with disciplinary differences on reflection. In other words, they believed that the design and application of evaluative metrics can be improved. They were more critical about the evaluation complacency held by administration and management because of the presumed objectivity of metrics or the simplicity for reporting and accountability. However, due to the nature of this study, the participants were likely to be more critical about metrics. Further studies would be required to understand whether academics in general are complacent about the use of evaluative metrics for research assessment and whether there are differences between disciplines.

The participants' everyday use of metrics in their research activities without resistance or questions can be described as evaluation inertia. Their use of the term 'game' or 'gaming' and their description of strategic behaviour to increase publication and citation counts signify their perception of not only the importance of evaluative metrics, but also how academia works. These practices are learned through day-to-day research activities and expectations set by universities and funding agencies. Whether evaluation inertia is effected by commercialisation and bureaucratisation of universities is beyond the scope of this paper; nevertheless, the participants' 'acceptance' of evaluative metrics in everyday research activities could be due to the apprehension that they are not doing good enough, fast enough, which leaves little time or intention to reflect and critique the legitimacy, objectivity, and rightfulness of existing standards—an inertia based on the perception of "how the system works" and its prescribed competitiveness.

Although there have been more open discussions about metrics in publications such as The Leiden Manifesto and The Metric Tide, responsible metrics is not a topic of urgency in everyday academic life, at least in the Irish context. The recruitment of participants in this study also provides some evidence of the lack of engagement in the discourse about metrics: considering the topic of metrics should be relevant to everyone, a much higher number of responses was expected. Although sufficiently rich data were collected at the end, the less than enthusiastic response rate, particularly in the Humanities, was a disappointment. The low engagement is a manifestation of the lack of discourse about metrics in Irish academic environment. Whether it is a matter of evaluation complacency or evaluation inertia demands further investigation.

With the assumption that productivity and quality are measured by the number of publications and citations in university rankings and so forth, the participants felt obliged to produce as much as they can. Strategies such as salami-slicing are deployed by some to game the system. It is also not uncommon for some participants to work on topics merely because the topics are regarded as trendy and fashionable by publishers and funding agencies. In the worst scenario, research outputs are framed and shaped by ready-made templates, like manufactured products churned out by a conveyer

belt. Out-of-the-box, truly novel and innovative ideas, and/or topics that take time can be discouraged, or simply not funded; or one may risk not being offered a position or promotion. Although the use of evaluative metrics is a motivation for some, it may deter others from investigating innovative topics or publishing in alternative publication outlets that do not have a high impact factor. Consequently, the creative, critical voices in academia could diminish over time. Evaluation complacency and evaluation inertia can be detrimental to knowledge production.

Open discussion about evaluative metrics would be beneficial for funding agencies, universities, publishers and other stakeholders to design better metrics and research evaluation mechanisms responsibly. These discussions should be based on good understanding of evaluative metrics, as well as a space for open and public discourse. The conflicting attitudes towards metrics in principle and in practice are but a symptom of malfunctioning system of knowledge production. Further studies can also investigate whether the conflicting attitudes are prevalent in contexts where the use of evaluative metrics is explicit and standardised.

6. Conclusion

This study shows that the use of metrics has been embedded in the participants' everyday research activities and academic life. The participants use metrics to track and evaluate their own and their peers' productivity and performance, and most consider metrics in their choices of publication channels, frequency, collaborators, as well as references to be included in their writings. Those in positions to assess recruitment, promotion, and funding proposals use metrics to make decisions. Together, these uses of metrics implicate that the objectivity, legitimacy, and rightfulness of metrics are regarded as generally accepted.

Their discussions about metrics during the interviews, however, reflect different opinions as to how evaluative metrics *should* be used. They articulated the importance of disciplinary differences in publication and citing practices, and more importantly, the importance of subject expertise in one's research field. There is a general agreement that metrics in and of themselves cannot represent the quality of research or the performance of an individual.

The discrepancy between what the participants think how evaluative metrics should be used and their own uses of metrics presents a conundrum, which has been explained by *evaluation inertia*, meaning the lack of engagement in the discourse about metrics in everyday research activities due to their perception of the importance of metrics and the necessity of productivity, on the one hand, and the seemingly non-existent platform for discourse about issues related to metrics, on the other. At the same time, the uses of evaluative metrics by administration and management units because of the simplicity of quantitative measures in reporting and quality assurance exercise, for example, are considered as a manifestation of *evaluation complacency*.

Nevertheless, research and publication practices are shaped by disciplinary norms and institutional requirements. The more evaluative metrics are used in the decision-making process of funding applications and academic appointments, the more they are normalised in everyday academic work. Although evaluative metrics does not have a one-to-one, cause-and-effect relationship with publication practices such as salami-slicing, the perception of their use as a standard of research

assessment have effects on the choice of research topics, publication channels, frequency of publication, and so forth.

The contents of academic publications are, by and large, products of academic labour. They are writings by academics and they are peer reviewed and edited by academics. At the same time, these products of academic labour are evaluated upon and ranked by evaluative metrics. Journal impact factors affect visibility, hence affect citation counts and h-index, amongst other indices. University rankings are calculated, in part, by aggregating citation counts. Evaluative metrics can hence be used to leverage the production cycle when incorporated in a system with an assumption that more is better, that is, the assumption that more publications or higher citation counts mean higher productivity and quality. Meanwhile, more products of evaluative metrics are being marketed to academics. Consequently, evaluative metrics are becoming more significant in the choice of research topics and publications by way of altering research practices and citing behaviour.

The most used evaluative metrics today are not necessarily the best indicators of the quality and the significance of academic work. Nevertheless, universities and academics not only passively accept these metrics, but also establish or conform to a system of reward using them. For universities, university rankings play a major role in the competitive higher education market in order to attract better, as well as foreign, students (Hammarfelt, de Rijcke, & Wouters, 2017); the rankings also serve the purpose of accountability, particularly when it comes to reporting. For academics, as many participants in this study articulated, the system encourages short-termism. Academics are running on the wheel of short cycles to publish, to get cited, to boast about their work on social media and the like. The validity of evaluative metrics and legitimacy of their use for evaluating research outputs, however, are rarely discussed or contested, but as a matter of passive acquiescence in everyday academic life. The findings of this study align with the discussion about the culture of speed in academy (Berg & Seeber, 2016, Vostal, 2015) and the changing ideas and ideals of universities (for example, Collini, 2012). Hence, the proliferation of publications is not necessarily a symptom of productivity or quality but a byproduct of the reward system effected by evaluative metrics.

Notwithstanding the quality and significance of publications, the perpetual push to publish leaves little time for academics to read, to reflect, to critique, and to review. It is not unreasonable to speculate that the quality of publications *on average* has been declining as the very system to ensure the integrity of research is not sustainable (see, for example, Newton, 2017), particularly when publications are counted while reviews are not formally recognized in the academic reward system.

Scientific discoveries can be accidental (e.g., penicillin, sildenafil). Out-of-the-box, unexpected results could be put aside if scientists were under immense pressure to publish and to be cited, and when producing failed but useful results were not 'counted.' It is also important to consider some of the most important scientific, philosophical and literary works were written many centuries ago and are still influential today. What does it mean if we were to attach a bundle of metrics to these giants? The Nobel Laureate in Physics, Peter Higgs, has told the mass media that "his lack of productivity probably would have gotten him fired long ago if he had not been nominated for a Nobel Prize" (Overbye, 2014). Evaluative metrics are not merely measures of research outputs, they also affect research practices, citing behavior and, in the end, knowledge production. As the commercial producers of evaluative metrics continues to emphasize that academics are obliged to include a basket of metrics on their research profiles, may universities, funding agencies, and academics reflect on the *actual benefits* of the use of evaluative metrics? As one participant noted that "*If it is not measuring*

what it's claimed to measure, then it's doing more harm than good." (Participant #6, Assistant Professor, Sciences).

To conclude, evaluative metrics should not be standardized and institutionalized without a thorough examination of their validity and reliability and without having their influences on academic life, research practices, and knowledge production investigated. There is also a need for 'metrics literacy', and, not the least, an open and public discourse about metrics. Conversations about metrics should not be limited to subject expertise and policy-makers but all who uses and is affected by metrics and related issues. Yet, evaluative metrics should not be considered as the sole factor contributing to research practices and academic life as they are embedded in the larger systems of research and national priorities. The structuration of research practices should be further studied taking into account the complexity of the academic lifeworld.

References

Aagaard, K. (2015) 'How incentives trickle down: Local use of a national bibliometric indicator system', *Science and Public Policy*, *42*:725–737.

Aksnes, D. W. and Rip, A. (2009) 'Researchers' perceptions of citations', *Research Policy*, 38/6: 895–905.

Amit, V. (2000) 'The university as panopticon: moral claims and attacks on academic freedom'. In M. Strathern (Ed.) *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*, pp. 215–235). London: Routledge.

Arnold, D. N. and Fowler, K. K. (2011) 'Nefarious numbers', *Notices of the AMS*, 58/3: 434–437.

ASCB (2013) *San Francisco Declaration for Research Assessment*. Retrieved June 15, 2017, from http://www.ascb.org/files/SFDeclarationFINAL.pdf

Berg, M. and Seeber, B. K. (2016) *The Slow Professor: Challenging the Culture of Speed in the Academy*. University of Toronto Press.

Blockmans, W., Engwall, L., and Weaire, D. (Eds.) (2014) *Bibliometrics: Use and abuse in the review of research performance*. Portland Press.

Blyth, E., Shardlow, S. M., Masson, H., Lyons, K., Shaw, I. and White, S. (2010) 'Measuring the quality of peer-reviewed publications in social work: Impact factors—liberation or liability?', *Social Work Education*, 29/2: 120–136.

Bornmann, L., and Daniel, H. (2008) 'What do citation counts measure? A review of studies on citing behavior', *Journal of Documentation*, 64/1: 45–80. http://doi.org/10.1108/00220410810844150

Bornmann, L. and Leydesdorff, L. (2014) 'Scientometrics in a changing research landscape: bibliometrics has become an integral part of research quality evaluation and has been changing the practice of research', *EMBO Reports*, 15/2: 1228–32.

Buela-Casal, G. and Zych, I. (2012) 'What do the scientists think about the impact factor?', *Scientometrics*, 92/2: 281–292.

Burrows, R. (2012) 'Living with the H-Index? Metric assemblages in the contemporary academy', *The Sociological Review*, 60/2: 355–372.

Butler, L. (2003) 'Modifying publication practices in response to funding formulas', *Research Evaluation*, 12/1: 39–46.

Butler, L. (2004) 'What happens when funding is linked to publication Counts?' In *Handbook of Quantitative Science and Technology Research*, pp. 389–405. Dordrecht: Kluwer Academic Publishers.

Clark, W. (2006) *Academic Charisma and the Origins of the Research University*. Chicago: The University of Chicago Press.

Clarke, C. A., & Knights, D. (2015) 'Careering through academia: Securing identities or engaging ethical subjectivities?', *Human Relations*, 68/2: 1865–1888.

Cleere, L. & Ma, L. (2018) 'A Local Adaptation in an Output-Based Research Support Scheme (OBRSS) at University College Dublin', *Journal of Data and Information Science*, 3/4: 73-83.

Collini, S. (2012) *What Are Universities For?* London: Penguin.

Craig, R., Amernic, J. and Tourish, D. (2014) 'Perverse audit culture and accountability of the modern public university', *Financial Accountability & Management*, 30/1: 1–24.

Cronin, B. (2005) *The Hand of Science: Academic Writing and Its Rewards*. Lanham, Maryland: Scarecrow.

Curry, S. (2012, January 15) 'Sick of impact factors', retrieved April 17, 2017, from http://occamstypewriter.org/scurry/2012/08/13/sick-of-impact-factors/

Dahler-Larsen, P. (2012) *The Evaluation Society*. Stanford, California: Stanford Business Books.

Dahler-Larsen, P. (2014) 'Constitutive Effects of Performance Indicators: Getting beyond unintended consequences', *Public Management Review*, 16/7: 969–986.

Day, R. E. (2014) "'The Data--It is *Me*!" ("Les données--c'est *Moi*!")' In B. Cronin & C. R. Sugimoto (Eds.) *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, pp. 67–84. Cambridge, Massachusetts: The MIT Press.

Davis, P. (2008) 'Elsevier Math Editor Controversy', retrieved June 13, 2018, from https://scholarlykitchen.sspnet.org/2008/11/25/elsevier-math-editor-controversy/

de Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P. and Hammarfelt, B. (2016) 'Evaluation practices and effects of indicator use—a literature review', *Research Evaluation*, 25/2: 161–169.

Erne, R. (2007) 'On the use and abuse of bibliometric performance indicators: a critique of Hix's "Global ranking of political science departments"', *European Political Science,* 6:304-314.

Franzoni, C., Scellato, G. and Stephan, P. (2011) 'Changing incentives to publish', *Science*, 333/6043. Available at http://science.sciencemag.org/content/333/6043/702.full

Furner, J. (2014) 'The ethics of evaluative bibliometrics'. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, pp. 85–108. Cambridge, Massachusetts: The MIT Press.

Garfield, E. (1955) 'Citation indexes for science', *Science,* 122: 108-111.

Garfield, E. (1972) 'Citation analysis as a tool in journal evaluation', *Science*, 178: 471-479.

Hammarfelt, B. and de Rijcke, S. (2015) 'Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University', *Research Evaluation*, 24: 63–77.

Hammarfelt, B., de Rijcke, S. and Wouters, P. (2017) 'From Eminent Men to Excellent Universities: University Rankings as Calculative Devices', *Minerva*, 55/4: 1–21.

Hicks, D. (2012) 'Performance-based university research funding systems', *Research Policy*, 41/2: 251–261.

Hicks, D. et al. (2015) 'The Leiden Manifesto for research metrics', *Nature*, 520: 429-431.

Johnson, M. H., Cohen, J. and Grudzinskas, G. (2012) 'The uses and abuses of bibliometrics', *Reproductive BioMedicine Online*. http://doi.org/10.1016/j.rbmo.2012.03.007

Kallio, K.-M., Kallio, T. J., Tienari, J. and Hyvonen, T. (2016) 'Ethos at stake: Performance management and academic work in universities', *Human Relations*, 69/3: 685–709.

Lamont, M. (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, Massachusetts: Harvard University Press.

Lariviere, V., Sugimoto, C.R. and Cronin, B. (2012) 'A bibliometric chronicling of Library and Information Science's first hundred years', *Journal of the American Society for Information Science & Technology,* 63/5: 997-1016.

Lawrence, P. A. (2007) 'The mismeasurement of science', *Current Biology*, 17/15: R583–5.

Liang, L., Zhong, Z. and Rousseau, R. (2014) 'Scientists' referencing (mis)behavior revealed by the dissemination network of referencing errors', *Scientometrics*, 101/3: 1973–1986.

MacRoberts, M. H. and MacRoberts, B. R. (1989) 'Problems of citation analysis: A critical review', *Journal of the American Society for Information Science*, 40/5: 342–349.

Markel, H. (2013) 'The real story behind penicillin', retrieved June 8, 2017, from http://www.pbs.org/newshour/rundown/the-real-story-behind-the-worlds-first-antibiotic/

Müller, R. and de Rijcke, S. (2017) 'Exploring the epistemic impacts of academic performance indicators in the life sciences', *Research Evaluation*, 26/3: 157–168.

Mutz, R. and Daniel, H.-D. (2012) 'Skewed citation distributions and bias factors: Solutions to two core problems with the journal impact factor', *Journal of Informetrics*, 6/2: 169–176.

Newton, A. (2017) 'The sustainability of peer review'. In *SpotOn Report: What Might Peer Review Look Like in 2030?* Retrieved June 7, 2017, from http://events.biomedcentral.com/wp-content/uploads/2017/04/SpotOn_Report_PeerReview-1.pdf

Ochsner, M., Hug, S. E. and Daniel H.-D. (Eds.) (2016) *Research assessment in the humanities*: *Towards criteria and procedures*. Springer Open.

Overbye, D. (2014) 'A Pioneer as Elusive as His Particle,' *The New York Times*. Retrieved June 8, 2017, from https://www.nytimes.com/2014/09/16/science/a-discoverer-as-elusive-as-his-particle-.html

Paradeise, C. and Thoenig, J.-C. (2013) 'Academic Institutions in search of quality: Local orders and global standards', *Organization Studies*, 34/2: 189–218.

Parker, L. (2011) 'University corporatisation: Driving redefinition', *Critical Perspectives on Accounting*, 22/4: 434–450.

Parker, M. and Jary, D. (1995) 'The McUniversity: Organization, management and academic subjectivity', *Organization*, 2/2*:* 319–338.

Porter, A. L. and Rafols, I. (2009) 'Is science becoming more interdisciplinary? Measuring and mapping six research fields over time', *Scientometrics*, 81: 719.

Power, M. (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

QS World University Rankings – Methodology. (n.d.). Retrieved June 15, 2016, from https://www.topuniversities.com/qs-world-university-rankings/methodology

Ranking Methodology of Academic Ranking of World Universities - 2016. (n.d.). Retrieved July 15, 2017, from http://www.shanghairanking.com/ARWU-Methodology-2016.html

Readings, B. (1996) *The University in Ruins*. Cambridge, Massachusetts: Harvard University Press.

Rushforth, A. and de Rijcke, S. (2015) 'Accounting for Impact? The Journal Impact Factor and the Making of Biomedical Research in the Netherlands', *Minerva,* 53: 117-139

Sosteric, M. (1999) 'Endowing mediocrity: Neoliberalism, information technology, and the decline of radical pedagogy', *Radical Pedagogy*, *1*(1). Retrieved from http://www.radicalpedagogy.org/radicalpedagogy/Endowing_Mediocrity__Neoliberalism,_Information_Technology,_and_the_Decline_of_Radical_Pedagogy.html

Shore, C. (2008) 'Audit culture and illiberal governance: Universities and the politics of accountability', *Anthropological Theory*, 8/3: 278–298.

Thelwall, M. and Kousha, K. (2016) 'ResearchGate articles: Age, discipline, audience size, and impact', *Journal of the American Society for Information Science and Technology,*68/2: 468-479.

Van Dalen, H. P. and Henkens, K. (2012) 'Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey', *Journal of the American Society for Information Science and Technology*, 63/7: 1282–1293.

Vostal, F. (2015) 'Academic life in the fast lane: The experience of time and speed in British academia', *Time & Society*, 24/1: 71–95.

Weingart, P. (2005) 'Impact of bibliometrics upon the science system: Inadvertent consequences?', *Scientometrics*, 62/1: 117–131.

White, H. D. and McCain, K. W. (1998) 'Visualizing a discipline: An author co-citation analysis of information science, 1972-1995', *Journal of the American Society for Information Science*, 49/4: 327-355.

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

Wittgenstein, L. (1958) *Philosophical Investigations*. Prentice Hall.

Wouters, P. (1999) *The Citation Culture* (Doctoral Thesis, University of Amsterdam, The Netherlands). Retrieved from http://garfield.library.upenn.edu/wouters/wouters.pdf

Zacharewicz, T., Lepori, B., Reale, E. and Jonkers, K. (2019) 'Performance-based research funding in EU Member States—a comparative assessment', *Science and Public Policy*, 46/1: 105-115.