Four Claims on Research Assessment and Metric Use in the Humanities

by Björn Hammarfelt

EDITOR'S SUMMARY

Bibliometric evaluation for research in the field of sciences can be a good way to assess the quality and factual basis of claims and can lead to more funding for authors and for research work. However, due to the more diverse fields covered, this type of evaluation is less effective in the world of humanities. Many professionals and researchers in humanities fields believe that bibliometric evaluation is meant only for STEM research and can't properly assess any findings made in humanities. Four common claims made about bibliometrics in humanities are that bibliometrics do not adequately cover the non-uniform nature of humanities; greater bibliometric coverage will not solve all the research problems in humanities subjects; metrics use already has an impact on humanities research practices and finally; other evaluation methods, like altmetrics, are conventional.

KEYWORDS

humanities
bibliometrics
evaluation
research and development
citation analysis
altmetrics

Björn Hammarfelt (Ph.D.) is a senior lecturer at the Swedish School of Library and Information Science (SSLIS), University of Borås, Borås, Sweden, and a visiting scholar at the Centre for Science and Technology Studies (CWTS) at Leiden University in Germany. His research is situated at the intersection between information science and sociology of science, with a focus on the organization, communication and evaluation of research. More information and contact details can be found at www.hb.se/en/Research/Researchers/Hammarfelt-Bjorn/.

A New Open Humanities

Evaluation is an intrinsic feature of modern societies. In a recent keynote, political scientist Peter Dahler-Larsen suggested that we find it easier to imagine aliens coming to earth than to imagine a society without evaluation. Evaluation in its many forms is a key, perhaps even a defining aspect, of academic knowledge production. What makes certain statements scientific or scholarly is to some degree dependent on their production and presentation; but what really distinguishes academic knowledge is the rigorous assessment, often through various forms of peer review, to which it is subjected. In this essay I focus on one very specific form of evaluation – the quantified assessment of research in the form of metrics based on publications, citations or social media mentions and how these measures impact or may impact the humanities.

Diverse publication patterns and dependence on local languages and contexts as well as specific referencing practices are distinctive features that have rendered bibliometric indicators less applicable in the humanities [1]. The difficulty of using bibliometric measures to evaluate research has resulted in attempts to create alternative systems of evaluation that look at new sources of attention data or that try to take the characteristics and the heterogeneity of research into account. Still, many scholars in the humanities and the social sciences remain skeptical towards bibliometric indicators. In this paper I discuss the potential that these measures have for capturing research performance in the humanities. The consequences of further quantification, both for knowledge production and academic culture, will also be emphasized.

The essay is structured around four claims on the use of bibliometrics for evaluating the humanities. These four claims serve as a way of summarizing key insights regarding the use of bibliometrics for the humanities and also

point to aspects which, at least partly, have been overlooked by previous research. The brief orientation given here should however not be seen as exhaustive, and the claims made are explorative rather than definitive. The broader implications of these statements are further discussed in the concluding section. However, before zooming in on the humanities I will give a brief overview of metric use and its consequences for knowledge production and research practices more generally.

Indicator Use and Effects of Research Evaluation

When discussing the proliferation of metrics in science, reference is often made to an *audit* or *evaluation society* [2] [3]. Factors that contribute to the proliferation of metrics are the commodification and commercialization of science and the emergence of new public management (NPM). According to Power, among others, new public management in many cases drove the construction of research evaluation systems. The ideal of NPM can be simply put as a "desire to replace the presumed inefficiency in hierarchical bureaucracy with the presumed efficiency of markets" [2, p. 43]. It is, therefore, not surprising that NPM is blamed for the introduction of market mechanisms in academic knowledge production. However, to understand the attractiveness of these measures – also in contexts where they are not mandated from above – we need to consider how the use of bibliometric indicators ties into disciplinary traditions of assessment. In fact researchers are deeply engaged in using and constructing indicators, and in some disciplines it might even be warranted to talk about citizen bibliometricians [4]. While researchers' engagement in constructing and using indicators may limit their ability to blame outsiders for unfair and unproductive evaluation procedures, this involvement also signals that there are opportunities for researchers themselves to actually influence assessment practices and evaluation systems.

The first empirical findings suggesting that the use of bibliometrics for evaluating research might influence knowledge production and publication patterns emerged some 15 years ago in Australia. In 2003 Linda Butler showed how an allocation system rewarding articles in international journals led to an increase in the number of publications but a drop in relative citation impact (compared to an international average) [5]. Today, research on this topic has grown, and studies of changes in publication patterns on a more general level have been supplemented with studies that look at effects on knowledge production more generally.

A recent review of metric use and its effects has identified four ways in which bibliometrics influence research:

1. Indicator use might result in strategic behavior and goal displacement. Hence, researchers might focus on work tasks that give the most points in the system rather than on doing a good job more generally. For example, the effort to produce more articles, but perhaps of less quality, in order to score well in a specific system might be seen as an attempt at such strategic gaming.
2. Many evaluation systems have been shown to be biased against interdisciplinarity, and, in particular, systems using journal rankings might lead to unfair assessment of interdisciplinary research.
3. The implementation of evaluation systems can lead to task reduction, where tasks rewarded in the system are prioritized. Thus, activities that are made invisible in these systems – for example, editing books or writing reviews – might eventually be abandoned.
4. The implementation of bibliometric evaluation has institutional effects. For example, a university might try to recruit highly cited researchers to gain positions in university rankings, or such transfers may be instigated in order to increase the institution's score in national evaluation systems [6].

The Use of Bibliometrics in the Humanities: Four Claims

Claim 1: The humanities are not uniform.

The claim that the humanities are a heterogeneous set of fields that cannot be discussed as a coherent whole is not controversial, and most bibliometric researchers would certainly agree with this statement. However, in the bibliometric literature, and admittedly also in some of my own work, the humanities quite often are discussed as a unified whole rather than as a disparate set of research fields. This practice is problematic, especially when the fields included in the definition of the humanities differ between contexts and the border to the social sciences is often fluid. Depending on the classification used, for example, gender studies, pedagogy, history or

TABLE 1. The humanities

EUROPEAN REFERENCE INDEX FOR THE HUMANITIES	*WEB OF SCIENCE* SUBJECT CATEGORIES	OECD FIELD CLASSIFICATION
Anthropology	Archaeology	History and Archaeology
Archaeology	Architecture	History (history of science under Philosophy, Ethics and Religion); Archeology
Art Architectural and Design History	Art	
Classical Studies	Asian Studies	
Gender Studies	Classics	Language and Literature
History	Dance	General language studies; Specific languages; General literature studies; Literary theory; Specific literatures; Linguistics
History and Philosophy of Science	Film, Radio, Television	
Linguistics	Folklore	
Literature	Humanities, multidisciplinary	
Musicology	Languages & Linguistics	Philosophy, Ethics and Religion
Pedagogical and Educational Research	Literary reviews	Philosophy; History and Philosophy of Science and Technology; Ethics; Theology; Religious studies
Philosophy	Literary criticism and theory	
Psychology	Literature	
Religious Studies	Literature, African, Australian, Canadian	
	Literature, American	Arts (arts, history of arts, performing arts, music)
	Literature, British Isles	
	Literature, German, Dutch, Scandinavian	Arts, Art history, Architectural design, Performing arts studies (Musicology, Theatre science, Dramaturgy); Folklore studies; Studies on Film, Radio and Television
	Literature, Roman	
	Literature, Slavic	
	Medieval & Renaissance Studies	
	Music	Oher humanities
	Philosophy	
	Poetry	
	Religion	
	Theatre	

anthropology may be defined as a humanities discipline or a social sciences discipline. Table 1 illustrates this point using the European Reference Index for the Humanities, the *Web of Science* subject categories and the OECD field classification.

A reason for the practice of discussing the humanities as a distinct entity is probably the previously mentioned focus on the otherness of the humanities when it comes to bibliometric evaluation. While discussing the humanities as a unified whole might be reasonable in a broader discussion – as done in this paper – it might also result in rather simplified statements regarding the application of bibliometric measures. It is indeed the case that citation analysis as an evaluation method is less applicable in many disciplines in the humanities. Still, some fields such as linguistics or philosophy, are organized in a way that may, at least to a limited extent, allow for the use of such methods. Consequently, the statement that "bibliometrics are not suitable for the humanities" can be questioned as it builds on a reductive and simplified definition of humanities research.

Claim 2: Greater coverage will not solve all problems.

Much research on bibliometric evaluation of the humanities points to the limitations of existing databases to adequately capture humanities research. The problem is that leading databases primarily index articles in English language journals, and this focus is the main reason why bibliometric evaluation in the humanities is less feasible than in other areas. A crucial step for solving these difficulties would be to include other types of sources, like monographs, book chapters and journals in languages other than English within the scope of these databases. The recent introduction of a Book Citation Index could also be viewed against this background.

While the limitations of bibliometric data (which also affects STEM fields) is a major issue for attempts to evaluate research using citation counts, it is not the only and perhaps not even the most important reason why citation analysis is less applicable in many humanities fields. Citation analysis demands that the intended audience is rather narrow, but the audience of humanities research is quite diverse and not easily demarcated. Nederhof distinguishes three major audiences: international scholars, national scholars and a lay

audience with professionals (for example journalists, librarians, archivists, etc.) being seen as a possible fourth audience [1]. Only the first audience – international scholars – is represented in major citation databases such as *Web of Science* and *Scopus*, and even for this group the coverage is low. While extending the databases might lead to greater coverage, important groups (the public and professionals) are still omitted.

The heterogeneous audience for the humanities suggests that researchers potentially have a broad reach, including an audience outside the academy, which means that in some humanities fields recognition from peers is not the only way of building reputation. This diversity gives scholars in the humanities a considerable degree of freedom when choosing research topics, but at the same time it limits the possibility of attracting citations. The rural organization of research also suggests that it may take considerable time for research in the humanities to gather citations – a window of up to 10 years has been suggested by Glänzel [7] – yet research might, on the other hand, remain relevant. Hence, temporal dimensions, which so far have largely been overlooked in studies of research evaluation, are key aspect to consider when scrutinizing assessment procedures.

Another reason why greater coverage will not automatically allow for measuring impact through citation is the referencing practices, why and how you cite, of many fields in the humanities. Not only do references to source materials (e.g., literary works or historical documents) make up a considerable amount of the citations in some fields, but these references are used for a variety of purposes, and contradictory or negative references are relatively common compared to STEM fields. Consequently, the diverse audience and specific referencing practice, as well as the overall intellectual organization of many fields in humanities, is the chief reason why citation-based evaluation is less usable. These matters will not automatically go away with greater coverage.

Claim 3: Metrics and indicator use already affects research practices in the humanities.

Numerous studies suggest that bibliometric methods are ill-suited for evaluating research in the humanities, but they are still often employed for assessing institutions or individuals. Even if not directly affected by evaluation systems or performance-based resource allocation, many scholars in the humanities feel targeted and biased against. In my study with de Rijcke, a historian was quite frank in his views on bibliometric measurement: "I know quite a lot about bibliometric evaluation but I ignore it. It is a crazy system developed for other disciplines than my own" [8, p. 73].

Moreover, other respondents in our study suggested that publication practices are changing – from books to articles – due to the implementation of bibliometric measurement. While empirical findings do not show a general trend toward journal publishing, a strong tendency toward publishing in English and an increase in peer-reviewed publications is evident, at least in a local context. Still, many evaluation systems accentuate the importance of publishing articles which may result in tensions between disciplinary quality standards and the criteria used for evaluation. A young literary scholar noted, "It's a problem that the status of monographs is very uneven – they definitely count as an advantage in my field, but not in funding and general academia. Thus, I have focused on writing articles to be on the safe side..."[8, p. 70].

Younger researchers might be more affected by bibliometric evaluation and other outside pressures as they have not yet secured a permanent position. Many perceive the focus on publication strategies by young scholars as negative for research in the humanities. However, this focus can also be seen as part of an intra-disciplinary debate between generations where the criticism voiced against publication strategies can be interpreted as a conflict between older traditions of publishing and new practices oriented towards an international audience.

While the findings presented above give some indication how humanities scholars react to bibliometric measurement it is still too early to grasp how disciplinary practices might change due to the implementation of these measures. Nonetheless it is clear that bibliometrics can play a role also in the humanities, and an on-going study that I am involved in suggests that about one third (32%) of all researchers in the humanities have used citations or rankings in assessing or promoting their work. Moreover, findings indicate that humanities scholars use a range of measures from well-known journal indicators and the h-index to emerging alternative measures such as

ResearchGate scores or views on *Academia.edu*. The uptake of measures based on usage statistics or social media mentions could support notions that so-called altmetric measures indeed provide a feasible alternative to more conventional indicators. However…

Claim 4: Alternative metrics are rather conventional.

During recent years altmetrics has been suggested as an alternative for studying impact outside established databases, and its usefulness for evaluating disciplines that are not easily covered by current methods has been highlighted. While the possibility to assess impact outside academic journals indexed in citation databases is a major improvement, as is the possibility to measure impact instantly (not having to wait for citations to accumulate), I suggest that many alternative metrics suffer from the same limitations as more conventional approaches. Hence, several altmetrics measures are still limited to evaluating journal articles, many of them tend to still focus on an academic audience, and the coverage of non-English sources is low.

In 2014 only 10% of all output from humanities scholars in Sweden was covered by altmetric services, and we still know quite little about what these indicators actually measure. Perhaps it is illustrative that according to one of the largest suppliers of altmetric data, *Altmetric.com*, the highest ranked article in the humanities in 2016 was titled "Revealing a 5000-y-old beer recipe in China" (www.altmetric.com/top100/2016/#subject=History+%26+Archaeology). It is still too early to discard altmetrics as one possible route for evaluating research in the humanities, and it appears that scholars in many fields do find these measures somewhat useful.

Concluding Reflections

Peer review will surely continue to be the main method for evaluating research in the humanities. In fields where it is possible we might see peer assessment being combined with bibliometric measures, but how this type of informed peer review will work in practice is still rather unclear. A combination of different types of indicators – bibliometric, altmetric and perhaps other types of measures suggested in research on different data sources – might also be a way forward.

When selecting and developing evaluation systems and indicators it is of great importance that humanities scholars take active part in discussions on research quality. By engaging in defining criteria for evaluation, researchers themselves can help to evade systems that do not correspond to how knowledge is produced and valued in a particular field. Discussions around quality can serve as a way of reflecting on and improving research practices. Furthermore, a bottom-up perspective on research assessment might eventually also "help society to better understand what SSH's [social sciences and humanities] contribution to solving major societal challenges can be" [9, p. 1].

Furthermore, it is also important to accentuate the importance that teaching has in many disciplines. Teaching and the forming of a well-educated populace or in Humboldtian terms, cultivated citizens, are fundamental objectives for scholarship in the humanities. A clear separation between the roles of research and education is in my view detrimental to scholarship, yet the current trend of quantified assessment seems to reinforce this separation in the humanities.

Finally, it is important to underline that although bibliometric methods often are inadequate for evaluating humanities scholarship I do not mean to suggest that researchers in these fields should avoid assessment more generally. On the contrary, critical evaluation in seminars, lengthy reviews and discussions are intrinsic parts of research practices in the humanities, and the valuation of arguments and texts is an ongoing activity. The reluctance to reduce "research quality" into a few comparable and computable numbers should perhaps be viewed in the light of this long tradition of critical assessment.

In the larger perspective of increasing mistrust in the numbers game, it might be that the sciences should learn from the humanities, and not the other way around. ■

Resources on following page

Resources Mentioned in the Article

[1] Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics, 66*(1), 81–100.

[2] Power, M. (1997). *The audit society: Rituals of verification.* Oxford: Oxford University Press.

[3] Dahler-Larsen, P. (2012) *The evaluation society*. Stanford: Stanford University Press.

[4] Hammarfelt, B. & Rushforth, A. (in press). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, https://doi.org/10.1093/reseval/rvx018, p. 1-12.

[5] Butler, L. (2003) Explaining Australia's increased share of ISI publications – The effects of a funding formula based on publication counts. *Research Policy*, 32(1), 143–55.

[6] De Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use – A literature review. *Research Evaluation, 25*(2), 161-169.

[7] Glänzel, W. (1996). A bibliometric approach to social sciences. National research performances in 6 selected social science areas, 1990-1992. *Scientometrics, 35*(3), 291-307.

[8] Hammarfelt, B., & 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*(1), 63–77.

[9] Ochsner, M., Hug, S. & Galleron, I. (2017). The future of research assessment in the humanities: Bottom-up assessment procedures. *Palgrave Communications*. 3:2017.20.