Journal Article

Do we really need BIBLIO-metrics to evaluate individual researchers?

Author(s):
Mutz, Rüdiger

ETH Library

Rüdiger Mutz

ETH Zürich

Do we really need BIBLIO-metrics to evaluate individual researchers?

Frequently, metrics are applied as quantitative measurements in order to evaluate research performance, particularly in the assessment of individual researchers with regard to appointment procedures for professorships, research fellowships, and tenure tracks. Thereby, the primary focus are bibliometric indicators, measurements that are derived from the quantitative analysis of documents, such as those included in bibliographic databases (e.g. *Web of Science* or *Scopus*). Yet, in my opinion, it is neglected that the classical approach of research evaluation, i.e. peer review, is often based on quantitative measurements too (e.g. rating scales). In light of the strong criticism of the peer review procedure, especially of its low reliability in terms of a lack of agreement between the reviewers' ratings, possible alternatives are discussed intensely. Under consideration are, for instance, the post-publication peer review procedure, in which peers evaluate an article after its publication, and, as mentioned, bibliometric indicators.

Benefits and problems of bibliometric data

I think the reason why bibliometric indicators are preferred is mainly due of three reasons: First, bibliometric indicators are considered to be objective, as they are devoid of subjective evaluations and made available by independent database providers (e.g. *Elsevier*, *Thomson Reuters*). Second, bibliometric data is transparent and not anonymous. Both the publications of researchers as well as their citing references, which ultimately are countable citations, can be identified and verified in bibliographic databases. Thus, in principle, results of bibliometric analyses can be replicated. Third, bibliometric indicators allow an immediate and concrete interpretation of research performance. The information that a researcher has published 10 publications and 5 of those publications have been cited over 100 times will probably be more and immediately indicative of the research performance of said scientist than for example a rating scale point of 3.4 on a 5-point rating scale, which is the result of the average of several referees' ratings of the quality of research.

However, bibliometric databases should not be mistaken to be perfect. Indeed, not all scientific journals are covered in such databases, the documents may include bibliographic errors, there are name ambiguities, the affiliations of institutions are often incorrect, and any database update can lead to changes within the whole bibliographic database (for example by adding or removing journals). Furthermore, *Goodhart's law* is valid for bibliometric indicators as well: An indicator that is used for evaluation will itself become a target of a optimization and is, therefore, no longer a suitable measurement for evaluation. For instance, it is possible to increase citations of a publication if authors frequently cite their own work (problem of self-citations). Self-citations may serve as an example: Author(s) can increase the number of their citations by frequently citing themselves.

h-index and percentiles

A classic example of a bibliometric indicator that can be used for the assessment of individual researchers is the *h*-index, in the way it is for example implemented on Scopus by default.

"A scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p - h)$ papers have $\leq h$ citations each." ([1], p. 16569). On first sight the *h*-index appears to be very appealing; it is easy to calculate and combines quantity (number of publications) with quality (citation impact) [2]. However, the *h*-index is not without fault (e.g. [3], p. 78). For instance, it is field-dependent. Because of higher citation levels in the field, a researcher in the life sciences can be expected to have a higher *h*-index than a scientist in social sciences. Older researchers have the advantage that they were able to publish more than younger researchers. The *h*-index can only increase and can be influenced through self-citations. What it is that constitutes a high *h*-index is unclear, as there is no benchmark for comparison. This critical remark does not mean, however, that I question the application of bibliometric indicators in general, but rather that I suggest a critical usage of such indicators. In that way, the *h*-index could be complemented through the employment of percentile ranks [4, 5]. A percentile rank quotes the percentage of articles in a scientific field that are at or below the citation score of a given article. If, for example, a journal article ranks within the top 10%, it belongs to the 10% of the highest cited articles in the scientific field in which it has been published.

Conclusions

In my view, the initial question indicated in the title, whether bibliometrics should indeed be used for the evaluation of researchers, can be answered with "yes". Bibliometric indicators should be applied as a supplement to, not as a substitute for, peer review

("informed peer review"). In my view, following *Bornmann* and *Marx* [4], *Wouters*, *Glänzel*, *Gläser*, and *Rafols* [6], the *Leiden Manifesto* [7], and the *San Francisco Declaration on Research Assessment* (DORA) [8], the following factors have to be taken into account with regard to the usage of bibliometric indicators:

1. Journal-based metrics, such as the journal impact factor, should no be used as a measure of the quality of individual research articles ([8] p. 869).

2. Research assessment should consider more than one bibliometric indicator with regard to the target of the evaluation. In particular, basic measures (number of publications, number of citations) and percentiles should be used ([8] p. 869, [7], [6] p. 50, [4])

3. Research assessment should analyse scientific collaboration patterns and subject profiles of individual researchers as well ([6] p. 50).

4. It is preferable to use publication lists authorised by the authors in order to conduct bibliometric analyses rather than to solely trust information provided by databases.

In my view, the usage of bibliometric indicators will strongly depend on whether the validity of those indicators with regard to external criteria (e.g. peer review ratings, scientific career, post publication peer review, ex-post evaluation of funded projects) can be proven. „*They need to be jointly tested and validated against what it is that they purport to measure and predict, with each metric weighted according to its contribution to their joint predictive power. The natural criterion against which to validate metrics is expert evaluation by peers* …" ([9] p. 103).

reporting correlations between the h index and 37 different h index variants. *J. Informetrics*, **2011**, *5*, 346–359. DOI: 10.1016/j.joi.2011.01.006

[3] Bornmann L, Marx W: The h index as a research performance indicator. *Eur. Sci. Ed.*, **2013**, *37*, 77–80.

[4] Bornmann L, Marx W: Evaluating individual researcher`s performance. *Eur. Sci. Ed.*, **2013**, 39, 39–40.

[5] Bornmann L, Mutz R: Further steps towards an ideal method of measuring citation performance: The avoidance of citation (ratio) averages in field-normalization. *J. Informetrics* **2011**, *5*, 228–230. DOI: 10.1016/j.joi.2010.10.009

[6] Wouters P, Glänzel W, Gläser J, Rafols I: The dilemmas of performance indicators of individual researchers. *ISSI Newsletter* **2013**, *9*, 48–53.

[7] Hicks D, Wouters P, Waltman L, de Rijks S, Rafols I: Bibliometrics: The Leiden Manifesto for research metrics. *Nature* **2015**, *520*, 429–431. DOI: 10.1038/520429a

[8] Cagan R: The San Francisco declaration of research assessment. *Dis. Model. Mech.* **2013**, *6*, 869–870. DOI: 10.1242/dmm.012955

[9] Harnard S: Validating research performance metrics against peer rankings. *Ethics Sci. Environ. Polit.* **2008**, *8*, 103–107.

Dr. Rüdiger Mutz
ETH Zürich
Professorship for Social
Psychology and Research
on Higher Education
Mühlegasse 21
8001 Zürich
Phone: +41 44 632 4918

mutz@gess.ethz.ch
www.psh.ethz.ch

References

[1] Hirsch JE: An index to quantify an individual's scientific research output. *PNAS* **2005**, *102*, 16569–16572, DOI: 10.1073/pnas.0507655102

[2] Bornmann L, Mutz R, Hug SE, Daniel HD: A multilevel meta-analysis of studies