CrossMark

Letter to the Editor: About the quality and impact of scientific articles

Martin Ricker[1] (iD)

Abstract It is argued that counting the total number of times a scientific article is cited by others, does neither result in a proxy for its cognitive impact nor for its quality. One would have to distinguish at least substitutable and fundamental references. A supposed correlation between peer review assessments and citation counts is conceptually problematic, because peer review includes objective as well as subjective considerations (convictions). With refined methods, however, a differential citation analysis might be able in the future to answer if a given article did or did not have positive cognitive impact on subsequent research.

Keywords Article impact · Article quality · Citation analysis · Peer review

Bornmann and Haunschild (2017) argue that by broadening the meaning of the term "impact" on societal impact, scientometrics is likely to lose its focus on quality, where "quality of publications was measured as a rule in terms of the number of citations" (p. 938). Furthermore they state that "impact of research might no longer be seen as a proxy for its quality" (p. 939). I like to comment on their assumption that

$$\text{citation count for article} \approx \text{its impact} \approx \text{its quality}. \tag{1}$$

Scientometrics, bibliometrics, informetrics, etc. emphasize their focus on mathematical and statistical analyses (see Milojević and Sugimoto 2012; Zhang et al. 2013). Statistics helps to analyze data logically, and to reveal patterns and conclusions that may not be immediately obvious, but ultimately you can always express in plain English what

✉ Martin Ricker
  mricker@ib.unam.mx

[1] Instituto de Biología, Universidad Nacional Autónoma de México (UNAM), Circuito Zona Deportiva, Ciudad Universitaria, 04510 Ciudad de México, Mexico

information is provided by the original data. Does citation count data contain the information about cognitive impact and article quality?

It is widely acknowledged that there are different motives for scientists to cite one another (Nicolaisen 2007). For example, Camacho-Miñano and Núñez-Nickel (2009: 757) distinguish nine categories, while Ricker (2015: 208) proposed to distinguish at least "substitutable citations" (called "perfunctory" in Waltman et al. 2013: 636) and "fundamental citations" (with cognitive impact). Substitutional citations are almost randomly chosen from a pool of thematically related articles, just to show that others have worked in the field. Most citations are substitutable (Alves Ramos et al. 2012; Ricker 2015: 208; Simkin and Roychowdhury 2005). Fundamental citations are the real target of evaluative citation analysis. If one lumps together a majority of substitutable citations and a remainder of fundamental citations in the data, one does not get an indicator of cognitive impact. The conclusion that citation counts are a proxy for cognitive impact becomes logically wrong. I am not aware of any evaluative scientometric analysis up to date, where the distinction has been made.

The term "quality" is even more difficult to connect with citation counts. To judge quality, one has to define the intrinsic characteristics that define it. In scientific articles, generally considered key characteristics are a significant contribution to current knowledge and expansion of the scientific frontier, creativity, clarity of exposition, and possibly applications. Time and again, the severe problem for the scientific process of using citation counts of scientists' articles as quality and impact parameters for evaluating the authors, has been pointed out (e.g., Adler and Harzing 2009; Allen et al. 2009; Belcher et al. 2016; Chavalarias 2017; De Bellis 2009: chapter 7; Gagolewski 2013; Hicks et al. 2015; Kaur et al. 2015; Ricker et al. 2009, 2010). Bornmann and Haunschild (2017) are aware of the problem, as they cite a number of studies (e.g., MacRoberts and MacRoberts 2010), and discuss the issue, also in previous own work (e.g., Bornmann et al. 2008). Their justification to go ahead with (1) is a supposed correlation between citation counts and peer review judgment (p. 938). To assume such a correlation is conceptually problematic, and would be the topic of a critical review by itself. Aksnes and Taxt (2004: 39) point out that a positive correlation between peer judgements and bibliometric performance measures can only be expected if the aspects assessed by the peers correspond to those reflected through bibliometric indicators. This, however, makes the argument largely circular. Furthermore, peers use generally a threshold approach to evaluation, after weighting all criteria: accept versus reject a manuscript, or interesting versus irrelevant publication for a given purpose. They rarely think about assigning a score on an open scale, like it is the case with citation counts.

Finally, and more fundamentally, peer evaluators make an informed, but partially subjective assessment, even when attempting to be as fair as possible. A justification is objective if in principle it can be tested and understood by anybody, while subjectivity refers to feelings of conviction (Popper 1959: 44, citing Immanuel Kant). Expert evaluations include objective and subjective elements. An "objective *evaluation*" is impossible, because *values* are subjective. Only an evaluation taking into account objective facts, such as scientometric indicators, is possible. The idea to automate evaluation completely with an objective algorithm, that has been discussed sometimes in scientometrics (e.g., Good et al. 2015; Nicolini et al. 1995), is as impossible as substituting a panel of supreme court judges with a computer. If one tells a computer to find the research applicant with the highest number of citations, then your subjective preference is to maximize the number of citations. One could have different preferences, such as finding a researcher who works with particular methods, or who works on the best applications for society (Hicks et al. 2015; Sutherland et al. 2011). Bibliometric indicators unfortunately often become "pseudo-

objective", when decision-makers use them as automated evaluation tools, with the necessary and unavoidable preference statements hidden behind them. The concept of preference statements of expert panels (see Ricker 2015), as implemented for the interface of producers and consumers in economics (and also econometrics), is notoriously missing in scientometrics!

One can come up with examples, where the goal of publishing a high-quality article is not related with a supposed goal to get widely cited. In my own institute, an important activity is the description of plant and animal species that are new to science. Assume that a new tree species is published in a taxonomic journal. Furthermore, assume that the article will receive in the future (maybe decades later) a single citation as a recognized species in a taxonomic revision. Then the mission of the original publication has been completed, with a single citation! It is all the cognitive impact you need to justify the publication, and the argument that it is a low-quality publication is wrong.

A good reflection for employing statistical models was formulated by the statistician George E. P. Box (1919–2013): "All [here scientometric] models are wrong, but some are useful" (Box et al. 2005: 384). Given the popularity that scientometrics and related fields have gained nowadays for assisting and often *de facto* replacing traditional research evaluation (e.g., Butler 2008; Good et al. 2015; Ricker 2015), scientists working in the field have also gained a lot of responsibility. Fortunately, there are some interesting developments in scientometrics, coming along with ever more advanced possibilities to analyze scientific texts with computers. First, it has been recognized that citation analysis can be carried out in much more detail: Which references are cited only once, and which ones more than once in the same text? One expects that recurrence of citations in a single article has to do with a higher importance for the citing author. In which sections are references cited? In the introduction, in the methods, or in the discussion? (see, e.g., Ding et al. 2013; Hu et al. 2015). The introduction tends to include substitutable citations, while citations in the methods or discussion sections are expected to indicate a more transcendental role, related to cognitive impact. Second, scientific text can be analyzed semantically in ever more detail: Which text segments have appeared in previous articles in similar form? How distant are semantically articles of the same authors from each other? How distant is a given article from other articles in the same scientific field? In which semantic context is an article cited? (see, e.g., Bertin et al. 2016; Gerrish and Blei 2010; Milojević 2015).

Easier than to quantify the impact of an article on a numerical scale would be to tackle the question if a given article did or did not have positive cognitive impact on subsequent research. With refined methods in the future, possibly a differential citation analysis will be able to answer this question automatically. Such a development I would call a "revolution in scientometrics", sensu Bornmann and Haunschild (2017).

References

Adler, N. J., & Harzing, A. W. (2009). When knowledge wins: Transcending the sense and nonsense of academic rankings. *Academy of Management Learning and Education, 8*(1), 72–95.

Aksnes, D. W., & Taxt, R. E. (2004). Peer review and bibliometric indicators: A comparative study at a Norwegian university. *Research Evaluation, 13*(1), 33–41.

Allen, L., Jones, C., Dolby, K., Lynn, D., & Walport, M. (2009). Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLoS ONE, 4*(6), e5910.

Alves Ramos, M., Gomes Melo, J., & Albuquerque, U. P. (2012). Citation behavior in popular scientific papers: What is behind obscure citations? The case of ethnobotany. *Scientometrics, 92,* 711–719.

Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R., & Zornes, D. A. (2016). Defining and assessing research quality in a transdisciplinary context. *Research Evaluation, 25,* 1–17.

Bertin, M., Atanassova, I., Sugimoto, V. L., & Lariviere, V. (2016). The linguistic patterns and rhetorical structure of citation context: An approach using n-grams. *Scientometrics, 109,* 1417–1434.

Bornmann, L., & Haunschild, R. (2017). Does evaluative scientometrics lose its main focus on scientific quality by the new orientation towards societal impact? *Scientometrics, 110,* 937–943.

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H. D. (2008). Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics, 8,* 93–102.

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters* (2nd ed.). Hoboken, NJ: Wiley.

Butler, L. (2008). Using a balanced approach to bibliometrics: Quantitative performance measures in the Australian Research Quality Framework. *Ethics in Science and Environmental Politics, 8*(83–92), 2008.

Camacho-Miñano, M. M., & Núñez-Nickel, M. (2009). The multilayered nature of reference selection. *Journal of the American Society for Information Science and Technology, 60,* 754–777.

Chavalarias, D. (2017). What's wrong with science? Modeling the collective discovery processes with the Nobel game. *Scientometrics, 110,* 481–503.

De Bellis, N. (2009). *Bibliometrics and citation analysis: From the science citation index to cybermetrics.* Lanham, MD: The Scarecrow Press.

Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics, 7,* 583–592.

Gagolewski, M. (2013). Scientific impact assessment cannot be fair. *Journal of Informetrics, 7,* 792–802.

Gerrish, S. M., & Blei, D. M. (2010). A language-based approach to measuring scholarly impact. In *Proceedings of the 27th international conference on machine learning*, Haifa, Israel.

Good, B., Vermeulen, N., Tiefenthaler, B., & Arnold, E. (2015). Counting quality? The Czech performance-based research funding system. *Research Evaluation, 24,* 91–105.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature, 520,* 429–431.

Hu, Z., Chen, C., & Liu, Z. (2015). The recurrence of citations within a scientific article. In *Proceedings of the 15th international society of scientometrics and informetrics conference*, Istanbul, Turkey (pp. 221–229).

Kaur, J., Ferrara, E., Menczer, F., Flammini, A., & Radicchi, F. (2015). Quality versus quantity in scientific impact. *Journal of Informetrics, 9,* 800–808.

MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology, 61*(1), 1–12.

Milojević, S. (2015). Quantifying the cognitive extent of science. *Journal of Informetrics, 9,* 962–973.

Milojević, S., & Sugimoto, C. R. (2012). Introduction: Metrics & ASIS&T. *Bulletin of the American Society for Information Science and Technology, 38*(6), 9–11.

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology, 41*(1), 609–641.

Nicolini, S., Vakula, S., Italo Balla, M., & Gandini, E. (1995). Can the assignment of university chairs be automated? *Scientometrics, 32*(2), 93–107.

Popper, K. R. (1959, reprinted 1992). *The logic of scientific discovery*. New York: Routledge.

Ricker, M. (2015). A numerical algorithm with preference statements to evaluate the performance of scientists. *Scientometrics, 103,* 191–212.

Ricker, M., Hernández, H. M., & Daly, D. C. (2009). Measuring scientists' performance: A view from organismal biologists. *Interciencia, 34,* 830–835.

Ricker, M., Hernández, H. M., & Daly, D. C. (2010). Contrasting views on Mexico's National System of Researchers: Answer by Ricker, Hernández and Daly. *Interciencia, 35*(3), 158–159.

Simkin, M. V., & Roychowdhury, V. P. (2005). Stochastic modeling of citation slips. *Scientometrics, 62,* 367–384.

Sutherland, W. J., Goulson, D., Potts, S. G., & Dicks, L. V. (2011). Quantifying the impact and relevance of scientific research. *PLoS ONE, 6*(11), e27537.

Waltman, L., van Eck, N. J., & Wouters, P. (2013). Counting publications and citations: Is more always better? *Journal of Informetrics, 7,* 635–641.

Zhang, L., Thijs, B., & Glänzel, W. (2013). What does scientometrics share with other "metrics" sciences? *Journal of the American Society for Information Science and Technology, 64*(7), 1515–1518.