

# Evaluations by Peer Review in Science

Lutz Bornmann

Received: 7 November 2012 / Revised: 13 December 2012 / Accepted: 18 December 2012  
© Springer International Publishing AG 2013

**Abstract** The development of modern science is closely connected to the assessment of research by peers. Evaluations by peers can be traced back until the middle of the 17th century. Nowadays, it is recommended for major evaluation contexts to use an informed peer review process where the peers are informed by bibliometric indicators. This review gives a short overview of the practice of peer review and its linking with bibliometrics.

**Keywords** Research evaluation · Peer review · Bibliometrics

## Evaluations in Science

Evaluations are an inevitable hallmark of science; scientists who are able to escape confrontation with ‘sceptical’ experts over their work are few and far between. ‘Quality evaluations are at the heart of the scientific enterprise. Such evaluations and their criteria and measurement determine who gets which job, who gets tenure, who gets which grants and who gets which awards and honours’ [17]. ‘Nearly every aspect of every role performed by scientists is evaluated with potentially consequential effects for scientists and their careers’ [25].

## Peer Review as a Tool for Evaluation

The first forms of peer review arose with the foundation of scientific societies—particularly the Royal Society in England—in the middle of the 17th century. They were intended to ensure the quality of the manuscripts which were published in the journals of the societies [5, 31, 35]. The societies were faced with the problem that people who were not members and therefore not seen as trustworthy by a society were presenting experimental results the quality of which was not at all clear. In order to render the results trustworthy, the societies either permitted a member to guarantee the correctness of the results or several members of the society undertook a critical review of the scientific work. Initially, the people had the experiments repeated under the members’ observation; later they reviewed the description of the process and the results of the experiments in manuscripts.

At first, the peer review procedure was applied exclusively to the review of manuscripts which were to appear in the scientific societies’ own journals; in modern science, it has developed into the most important evaluation tool for scientific work and is used to review planned (ex ante) and completed (ex post) research [19, 28]. The procedure is not only used to select manuscripts for publication in journals [20]; it also determines who is awarded prizes (such as the Nobel Prize), grants and jobs. Furthermore, it is used to evaluate research groups and scientific institutions [24], whereby not only research might be under review, but also the scientists’ teaching activities (evaluation of study and teaching), see [8]. With universities forced to make savings, a trend has become apparent over the years in which scientists have less access to normal funding from their universities and increasingly must finance their research from external funding which is granted with the peer review process [21, 29].

---

L. Bornmann (✉)  
Division for Science and Innovation Studies, Administrative  
Headquarters of the Max Planck Society, Hofgartenstr. 8,  
80539 Munich, Germany  
e-mail: bornmann@gv.mpg.de

As a rule, peer review processes are structured so that scientists are asked by an institution to review and legitimise an evaluation object and their expert report is submitted to decision-makers for further use (for example, to accept or reject a manuscript or an application for research). Depending on the area of application (and also within an area of application), the actual arrangements for the evaluation process can vary to a greater or lesser degree [23, 30, 40]. The evaluation can be undertaken in accordance with highly formalised specifications, but the selection of criteria on which to judge the work can also be left to the referees. Referees can remain anonymous or be named. The researchers under review can be anonymous or their names can be visible to the referees ('double-blind' vs. 'single-blind'). A referee can be deployed on a permanent or an ad hoc basis [19]. An evaluation can be carried out by one referee or by several [30]. If a group of referees is used, the members can evaluate the work independently or work with the others [19]. The peer review process can be designed to be accessible to the public or it can be confidential.

Typically, peer review is used in science to evaluate manuscripts and research applications. In these areas, the referees are tasked with recommending the 'best' scientific research given the scarce resources (that is, a limited number of printed pages in journals or limited resources for research) and possibly with formulating suggestions for improvement to the submitted work [15, 22]. They are also supposed to find errors in scientific work or incorrect scientific conduct in association with the work [41]. 'Peer review should be a filter to separate the wheat from the chaff, and thus to guarantee and improve the quality of published research' [20].

The proponents of peer review consider the process for checking and legitimising scientific work (manuscripts or applications) to be more suitable than any other proposed up to now. For example, Roy [37] developed a 'Peer Reviewed Formula System', in which research funding was granted in proportion to previous scientific productivity. However, this did not prevail as an alternative to peer review. A proponent of peer review, Abelson [2], wrote that 'the most important and effective mechanism for attaining good standards of quality in journals is the peer review system' (p. 62). Many years later, ScholarOne [39] expressed a similarly positive view: 'The peer review process is considered an integral part of scholarly communication, helping to ensure validity, increase accuracy, and enhance content prior to publication and dissemination'. A global survey on the attitudes and behaviour of 3,040 academics in relation to peer review in journals shows that 'the overwhelming majority (93 %) disagree that peer review is unnecessary. The large majority (85 %) agreed that peer review greatly helps scientific communication and most (83 %) believe that without peer review there would be no control' [34, p. 1].

The proponents of peer review consider active scientists to be the most suitable people to evaluate the scientific quality of work done by colleagues in their own field [14, 19]. In their view, peer review guarantees like no other procedure an evaluation of performance that is appropriate for the scientific requirements using the scientific expertise constitutionally required. 'When the peer review process works, statements and opinions are not arbitrary, experiments and data meet certain standards, results follow logically from the data, merit rather than influence determines what is published, and researchers do not have to waste their time reading unfiltered material' [32, p. 64]. The situation whereby scientists are both referees and refereed in the peer review process contributes to standardising the evaluation criteria and also to generalising the formal requirements [42].

Critics of peer review (see, for example, [1, 18, 26, 36, 37]) see a number of weaknesses in the process: (1) Different referees do not agree on the assessment of the same scientific work (lack of reliability). (2) Recommendations from referees are systematically biased. The work is judged not only on its scientific quality, but also on non-scientific criteria (lack of fairness). (3) Accordingly, the correlation between the judgements in the peer review process and the quality of the refereed work is low (lack of validity). The only reason, according to the critics, for continuing to use the peer review process is that there is no clear consensus on a 'better' alternative [51].

Research into the peer review process which has looked at the criticisms levelled against it relates largely to peer review for journals (see overviews in [4, 6, 9, 10, 33, 44, 49]) and less often to peer review of research and grant applications (see overviews in [6, 13, 50]). This research is criticised on the grounds that 'most of the publications on journal peer review are more opinion than research, often the ruminations of a former editor. Likewise, most of the letters to editors on the topic, the comments of one kind or another are predominantly opinion' [44, p. 215]. For example, when Overbeke and Wager [33] researched a number of peer review studies in the area of biomedicine, they found that many of these studies had been published as editorials, comments or letters and not as original research articles (see also [43]). An overview of research into the peer review process [12] summarises: 'While peer review is central to our science, concerns do exist. Despite its importance, it is curious that we have not required the same rigour of study of the peer review process as we do for our science' (p. 275).

### The Connection Between Bibliometric Indicators and Peer Review

While some see 'qualitative' peer review and 'quantitative' techniques of bibliometrics as competing approaches [7],

most others recommend that they should be complementary so that each compensates for the weaknesses of the other [46]. ‘Winking at the tradition of library studies, the term “bibliometrics”, coined by Alan Pritchard in the late 1960s, stresses the material aspect of the undertaking: counting books, articles, publications, citations, in general any statistically significant manifestation of recorded information, regardless of disciplinary bounds. “Scientometrics”, instead, emphasizes the measurement of a specific type of information, upon which a certain kind of value judgment—relative to the status of “scientific”—has already been made by someone put in charge of (and trusted for) giving it’ [11, p. 3].

On one hand, goes the argument, peer review is (1) informed by bibliometric indicators and also (2) subjected to verification. (1) ‘Bibliometric analysis will never be a substitute for peer review, but, if the analysis is comprehensive and sound, it should inform peer review’ [27, p. 321]. ‘The experience at INSERM [The French National Institute of Health and Medical Research, Paris] shows that introducing bibliometric indicators ... strengthens the transparency of decisions and provides for their substantiation’ [16]. (2) For Weingart [47], the indicators have an important function in that they subject the evaluations from a peer review to verification and thus protect them from the ‘old boys’ network effect’—a fundamental problem which afflicts the process [48].

On the other hand, the results of bibliometric analyses can be commented on and interpreted by referees. Schneider [38] considers that direct evaluation of scientific performance in accordance with formalised and context-independent criteria and without additional interpretation is impossible. It would require expert referees to undertake a relevance specification appropriate to the situation, taking into account the repercussions on individuals and institutions [47]. Only interpretation by experts would make it possible to develop qualitative assessments of scientific performance from the results of bibliometric analyses. The indicators would not only require annotation by an expert, but also enhancement and correction from an awareness of the context.

For Abramo and D’Angelo [3], the combination of peer review and bibliometrics offers a number of benefits: ‘The use of both methods, with the peer-reviewer having access to bibliometric indicators (hopefully appropriately standardised) concerning the publications to be evaluated, permits the reviewer to form an evaluation that emerges from comparison between his/her personal subjective judgment and the quantitative indicators. The pros and cons of the two approaches for evaluation of single scientific products are probably balanced, making it difficult to establish which would be preferable: the variables of context and the objectives of the evaluation could shift the weight in favour

of one or the other. In fact, it is not an accident that many studies have demonstrated a positive correlation between peer quality esteem and frequency of citation’ (p. 503).

## Conclusions

Nowadays, referees, editors, and programme managers are facing significantly increased demands. Whereas in the past their task was to filter out reliably that which did not meet a certain minimum standard (negative selection), today they have to choose the best from many, frequently very good, research applications (positive selection). Where the bodies who fund research are concerned, positive selection, unlike negative selection, is associated with a lower likelihood of supporting research projects (or people) which or who are later revealed to be less than successful. Where the applicants are concerned, however, positive selection can lead to the disappointing experience of having the funding for their project declined, although it later turns out, if they have acquired funding elsewhere, to be just as successful as funded projects [45].

There are good reasons to see ‘qualitative’ peer review and the ‘quantitative’ techniques of bibliometrics not as competing processes in scientific evaluation, but as two options with which to view a scientific work from different perspectives. It is only by linking both perspectives that a comprehensive view of the whole emerges, making possible *credible* statements about the quality of a scientific study.

## References

1. Abate T (1995) What’s the verdict on peer review? *Ethics Res* 1:1
2. Abelson PH (1980) Scientific communication. *Science* 209 (4452):60–62
3. Abramo G, D’Angelo C (2011) Evaluating research: from informed peer review to bibliometrics. *Scientometrics* 87(3):499–514. doi:[10.1007/s11192-011-0352-7](https://doi.org/10.1007/s11192-011-0352-7)
4. Armstrong JS (1997) Peer review for journals: evidence on quality control, fairness, and innovation. *Sci Eng Ethics* 3(1): 63–84
5. Biagioli M (2002) From book censorship to academic peer review. *Emergences* 12(1):11–45
6. Bornmann L (2011) Scientific peer review. *Annu Rev Inf Sci Technol* 45:199–245
7. Bozeman B, Dietz JS, Gaughan M (2001) Scientific and technical human capital: an alternative model for research evaluation. *Int J Technol Manag* 22(7–8):716–740
8. Brennan J, El-Khawas E, Shah T (1994) Peer review and the assessment of higher education quality: an international perspective (report no. 3). Open University Quality Support Centre, London
9. Campanario JM (1998) Peer review for journals as it stands today—part 1. *Sci Commun* 19(3):181–211

10. Campanario JM (1998) Peer review for journals as it stands today—part 2. *Sci Commun* 19(4):277–306
11. de Bellis N (2009) Bibliometrics and citation analysis: from the science citation index to cybermetrics. Scarecrow Press, Lanham
12. de Vries DR, Marschall EA, Stein RA (2009) Exploring the peer review process: What is it, does it work, and can it be improved? *Fisheries* 34(6):270–279
13. Demicheli V, Pietranonj C (2004) Peer review for improving the quality of grant applications (Cochrane methodology review). The Cochrane library, issue 1. Wiley, Chichester
14. Eisenhart M (2002) The paradox of peer review: Admitting too much or allowing too little? *Res Sci Educ* 32(2):241–255
15. Enserink M (2001) Peer review and quality: A dubious connection? *Science* 293(5538):2187–2188
16. European Science Foundation (2007) Peer review. Its present and future state (conference report). European Science Foundation (ESF), Strasbourg
17. Feist GJ (2006) The psychology of science and the origins of the scientific mind. Yale University Press, New Haven
18. Finn CE (2002) The limits of peer review. *Educ Week* 21(34):30–34
19. Geisler E (2000) The metrics of science and technology. Quorum Books, Westport
20. Grivell L (2006) Through a glass darkly: the present and the future of editorial peer review. *EMBO Rep* 7(6):567–570
21. Guston DH (2003) The expanding role of peer review processes in the United States. In: Shapira P, Kuhlmann S (eds) Learning from science and technology policy evaluation. Experiences from the United States and Europe. Edward Elgar, Cheltenham, pp 81–97
22. Hackett EJ, Chubin DE (2003) Peer review for the 21st century: applications to education research. Paper presented at the peer review of education research grant applications. Implications, considerations, and future directions. Washington, DC
23. Hansson F (2002) How to evaluate and select new scientific knowledge? Taking the social dimension seriously in the evaluation of research quality. *VEST* 15(2–3):27–52
24. Hemlin S (1996) Research on research evaluations. *Soc Epistemol* 10(2):209–250
25. Hermanowicz JC (2006) What does it take to be successful? *Sci Technol Hum Values* 31(2):135–152. doi:[10.1177/0162243905283637](https://doi.org/10.1177/0162243905283637)
26. Horrobin DF (2001) Something rotten at the core of science? *Trends Pharmacol Sci* 22(2):51–52
27. Jones P, Sizer J (1990) The universities funding council's 1989 research selectivity exercise. *Beiträge zur Hochschulforschung* 4:309–348
28. Langfeldt L (2006) The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. *Res Eval* 15(1):31–41
29. Laudel G (2006) The quality myth: promoting and hindering conditions for acquiring research funds. *High Educ* 52(3):375–403
30. Marsh HW, Ball S (1991) Reflections on the peer review process. *Behav Brain Sci* 14(1):157–158
31. McClellan JE (2003) Specialist control: the publications committee of the Académie Royale des Sciences (Paris), (Transactions of the American Philosophical Society), vol 93. American Philosophical Society, Philadelphia, pp 1700–1793
32. McCormack N (2009) Peer review and legal publishing: what law librarians need to know about open, single-blind, and double-blind reviewing. *Law Libr J* 101(1):59–70
33. Overbeke J, Wager E (2003) The state of the evidence: what we know and what we don't know about journal peer review. In: Godlee F, Jefferson T (eds) Peer review in health sciences, 2nd edn. BMJ Books, London, pp 45–61
34. Publishing Research Consortium (2008) Peer review in scholarly journals: perspective of the scholarly community—an international study. Publishing Research Consortium, Bristol
35. Resnik DB (2011) A troubled tradition it's time to rebuild trust among authors, editors and peer reviewers. *Am Sci* 99(1):24–27
36. Ross PF (1980) The sciences' self-management: manuscript refereeing, peer review, and goals in science. The Ross Company, Lincoln
37. Roy R (1985) Funding science: the real defects of peer-review and an alternative to it. *Sci Technol Hum Values* 52:73–81
38. Schneider WL (1988) Grenzen der Standardisierbarkeit bei der Bewertung von Forschungsergebnissen. In: Daniel H-D, Firsch R (eds) Evaluation von Forschung: Methoden, Ergebnisse, Stellungnahmen. Universitätsverlag Konstanz, Konstanz, pp 433–447
39. ScholarOne (2011) Increasing the quality and timeliness of scholarly peer review: a report for scholarly publishers. Thomson Reuters, Philadelphia
40. Shashok K (2005) Standardization vs diversity: How can we push peer review research forward? *Med Gen Med* 7(1):11
41. Smith R (2006) Peer review: a flawed process at the heart of science and journals. *J R Soc Med* 99(4):178–182
42. Souder L (2011) The ethics of scholarly peer review: a review of the literature. *Learn Publ* 24(1):55–72. doi:[10.1087/20110109](https://doi.org/10.1087/20110109)
43. Squazzoni F, Takacs K (2011) Social simulation that 'peers into peer review'. *JASSS* 14(4):3
44. Stieg Dalton MF (1995) Refereeing of scholarly works for primary publishing. *Annu Rev Inf Sci Technol* 30:213–250
45. Strange P (2007) Not so good when 75 % of grant applications fail. *Nature* 448(7149):22
46. van Raan AFJ (1999) Advanced bibliometric methods for the evaluation of universities. *Scientometrics* 45(3):417–423
47. Weingart P (2005) Das Ritual der Evaluierung und die Verführbarkeit. In: Weingart P (ed) Die wissenschaft der Öffentlichkeit: essays zum verhältnis von wissenschaft, medien und öffentlichkeit. Velbrück, Weilerswist, pp 102–122
48. Weingart P (2005) Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics* 62(1):117–131
49. Weller AC (2002) Editorial peer review: its strengths and weaknesses. Information Today, Inc., Medford
50. Wessely S (1998) Peer review of grant applications: What do we know? *Lancet* 352(9124):301–305
51. Young SN (2003) Peer review of manuscripts: theory and practice. *J Psychiatr Neurosci* 28(5):327–330