

# Peer Review and Bibliometrics

## 1 Abstract

Peer review is an established process supporting decisions made on journal publications, grant applications, and tenure, but also helping to assess research groups. For several reasons, peer review is currently being debated, and bibliometrics could serve as its substitute. A large number of studies comparing both approaches has been published. An overview of their results is presented and discussed in this chapter. Although there are good reasons to be hesitant about utilizing bibliometric approaches to assess single persons (e.g. for tenure), the situation is different when assessing research groups. In the STM area, bibliometric indicators could be used as a replacement for peer review.

## 2 Introduction

Peer review is an evaluation of a manuscript or a research proposal, for example. It is carried out by one or more people with expertise similar to that of the producers of the work, i.e. peers (Weller 2001). The term also includes the retrospective evaluation of the past performance of a scientist or a group of scientists (Gemma 2017). In a slightly broader sense, the term 'peer review' is also used to assess, for example, teaching (Chism and Chism 2007; Samson and McCrea 2008; Snively and Dewald 2011; Van Valey 2011) or medical staff and practices (Chop and Eberlein-Gonska 2012; Ertel and Aldridge 1977; Hadian et al. 2018; Lang 1999). The latter forms are not discussed here. This chapter opens with an overview of peer review, before presenting a selection of studies comparing peer review and bibliometric indicators. It concludes with a discussion of the findings.

## 3 Peer review

### 3.1 The development of peer review

It is often claimed – probably for the first time by Zuckerman and Merton (1971) – that peer review was invented by the first secretary of the Royal Society, Henry Oldenburg (1610-1677). Further occurrences of this narrative can be found in Baldwin (2018). This has led to the impression that peer review is more or less inextricably linked with scholarly publishing. However, Melinda Baldwin showed that it was only over the course of the nineteenth and early twentieth centuries that a number of learned societies adopted the practice of systematically consulting anonymous reviewers about submitted papers. For instance, the Royal Society of Chemistry adopted reviewing systems in the nineteenth century. The American Physical Society implemented this practice in the early twentieth century, though it was not until the 1960s that all submissions to their flagship journal *Physical Review* were peer-reviewed. At this time, the editors of *Nature* still abstained from consulting external reviewers if papers were submitted or recommended by scientists whom they trusted. It was not until 1973 that external peer review became mandatory for manuscripts submitted to *Nature* (Baldwin 2015). As late as 1989, an editorial in *The Lancet* revealed a huge inner distance to peer review: "In the United States far too much is being demanded of peer review. Careers and the viability of whole departments now depend on publication in peer-reviewed journals. In the public domain the process is sometimes seen as a guarantee of truth, which is silly; (...) Journals do things differently, and long live those differences, but there was consensus that turning away papers within the editorial board or 'in house' without an outside opinion by no means disqualified a journal from calling itself peer reviewed and that reviewers are advisers (always *The Lancet's* preferred term) not decision makers." (Anonymous 1989).

### 3.2 The peer-review process

Probably the most important type of peer review occurs when a manuscript is submitted to a journal (Paltridge 2017). Journals that have implemented a peer review process are referred to as 'peer-reviewed journals' or 'refereed journals'; sometimes only publications in such journals are considered 'real' scientific output. The first step in the process after submission is assessment by the editor. At this stage, some submissions are rejected ('desk rejection') because they are of low quality or because they are beyond the journal's scope. Once they pass this initial screening, articles are sent to external reviewers. In some cases, authors are invited to suggest reviewers. The number of reviewers ranges from one to three or even more at the discretion of the editor or depending on the regulations of the journal. The standard procedure is 'blind peer review', where the author receives the reviewers' comments, but the reviewers themselves remain anonymous. In 'double-blind peer review', the author is also anonymized. However, this is not always an easy task, for instance if information about the author's affiliation is necessary to understand the article. The rationale here is to enable reviewers to express their opinions openly and, in the case of double-blind peer review, to avoid any bias with regard to the authors' gender, age, reputation, etc. 'Open peer review' is the opposite approach, applied for example by journals published by Copernicus. Here, manuscripts are published in their initial form and then undergo a public review (everybody can comment) in addition to invited and more in-depth reviews. All comments can be read free of charge.

In most cases, reviewers are asked to express their opinion in a structured manner. The intention is to help the editor in judging the comments and the authors in implementing them. Most often, reviewers have to opt for one final recommendation; for example, 'accept as is', 'accept with minor revisions', 'accept with major revisions', 'reject'. If – in the case of two reviewers – there is a substantial disagreement between the recommendations, editors either make the final decision themselves or invite a third reviewer.

If a revision is deemed necessary, authors are advised to follow the reviewers' suggestions though it is possible in principle for the authors to discuss changes they may consider unreasonable with the editor. It is common practice for authors to provide a detailed response to each reviewer, explaining how they have followed the reviewer's advice or, if they have not, their reasons for not implementing certain suggestions. After revision and resubmission, the editor again decides whether the article will be published or not. Overall, rejection rates differ hugely between journals, e.g. between 2 per cent and 68 per cent in the atmospheric sciences (Schultz 2010).

The peer review process for books is rather heterogeneous (Goldfinch and Yamamoto 2012). External peer review by one reviewer (or even more than one) is often substituted by an editorial review, particularly in the case of edited books, where a number of different authors write the chapters. In the case of monographs and particularly in the case of text books, the publisher provides the editing through his staff.<sup>1</sup>

Grant proposals can be reviewed in exactly the same way as journal articles with the exception that here reviewers are not asked to improve the text of the proposal but rather to suggest a different experimental setting, etc. Often, grant proposals are judged by a review panel in quite a different setting: the reviewers do not act independently of each other but discuss the proposal together after questioning the applicants. This process often comprises a single session in which a number of proposals for the same tender are reviewed. As a grant application not only includes the research proposal itself, but also curriculum vitae of the participating scientists along with their publication

---

<sup>1</sup> In contrast to this, journal editors are usually not staff members of the publisher, but scientists who pursue this work in addition to their research and teaching duties. They get no or only minimal compensation for their editorial duties (de Knecht 2019).

records and letters of recommendations, the review also takes the past performance of the applicants into account.

An even greater shift towards past performance assessment occurs in the evaluation of research groups or institutions. For instance, in Germany, universities participating in the German federal and state governments' Excellence Initiative undergo a rigorous assessment carried out by scientists from abroad. The same applies to the extramural research institutions of the Leibniz Association and the Helmholtz Association. They are evaluated every seven years to decide whether they can continue as a member of the Leibniz Association or, in the case of Helmholtz, to what extent the research programmes will be financed. I will elaborate using the example of the Helmholtz Association: Helmholtz pools its research activities in six strategic programmes, which typically extend across the Association's research centres. Programme evaluation happens in two stages. First, each centre's past performance is evaluated in relation to the programme. Two years later, the proposals for future research are assessed. On average, each programme has a publication output of 20,000 journal articles in a seven-year period (Helmholtz 2018). The key questions here are: what is the most suitable method for assessing such a volume of scientific output and is such an assessment even feasible?

### 3.3 Peer review under criticism

Peer review is extremely important in scholarly communication, and there are a number of supporting arguments in its favour. First, there is the (mistaken) assumption that peer review has always been linked to scholarly publishing. Furthermore, article quality is considered to be improved by the process (Jefferson, Wager, and Davidoff 2002; Pierie, Walvoort, and Overbeke 1996). Finally, peer review is regarded as a form of self-regulation within a field: author and reviewer are on an equal footing and could theoretically swap roles. This distinguishes peer review from the examination of a thesis by a supervisor and from the application process in an office.

Nevertheless, criticism of peer review has been growing in recent years (L. Bornmann 2011; Eysenck and Eysenck 1992; Gould 2013; Smith 2006). Some of the criticized aspects include:

- (1) Peer review is inefficient, time consuming, and expensive. Publications are often delayed, and innovative and unconventional ideas can be prohibited.
- (2) Reviewers might deliberately reject the paper or at least slow down the process and in the meantime steal and publish the results on their own (Retraction Watch 2016).
- (3) A plurality of reviewers rarely agree on their recommendations so reproducibility and reliability is poor (Lutz Bornmann, Mutz, and Daniel 2010).
- (4) The fate of a particular proposal is only partially determined by its scientific value; random elements appear to play an important role as well (Cole, JR, and Simon 1981).
- (5) Reviewers' recommendations are often biased, which means that the fairness of the process is questionable.

## 4 Peer review and bibliometrics

In light of this critique, alternatives to peer review would be worth testing, even though the aforementioned approach of open peer review could overcome a number of problems associated with peer review (Godlee 2002; Walsh et al. 2000). In the case of grant applications, a system known as the 'peer-reviewed-productivity formula system' (Roy 1985) has been proposed, but it apparently has never been adopted. Another option is to utilize bibliometric indicators instead of peer review. This approach has been the subject of many investigations (for overviews, compare e.g. (Aksnes, Langfeldt, and Wouters 2019; Vieira and Gomes 2018; Gallo and Glisson 2018; Wouters et al. 2015; L. Bornmann 2011). For historical reasons (Wouters et al. 2015), the majority of these studies aim to

compare results derived from bibliometric indicators with peer review, thus taking peer review as the gold standard. However, the subsequent overview starts with studies which test peer review using bibliometric parameters.

#### 4.1 Peer review vs. Journal Impact Factor

It is commonly accepted that there is a hierarchy of journals. This is sometimes conveyed using the image of a pyramid: There are a few high-ranking journals (defined as such by their Journal Impact Factor (JIF), according to the Journal Citation Reports published by Clarivate Analytics), followed by a larger number of middle-tiered journals and an even larger number of low-impact journals (McKiernan et al. 2019). "To succeed in science, one must climb this pyramid: in academia at least, publication in the more prestigious journals is the key to professional advancement." (Jennings 2006). According to this assumption, authors will submit their manuscript to one of the highest ranked journals and in the case of a rejection move down the pyramid and submit to journals with a lower JIF. (L. Bornmann 2011) cites a dozen studies that have compared the quality of the rejecting and the subsequent accepting journals by means of the JIF. Up to 70 per cent of the rejected manuscripts could be tracked as having been published later in a journal with a higher JIF. Hence, "authors do not necessarily move from 'leading' journals to less prestigious journals after a rejection" (Weller 2001).

#### 4.2 Peer review vs. citation analysis

Sir Theodore F. Fox, former editor of *The Lancet*, was rather sceptical about the predictive value of editorial decisions: "When I divide the week's contributions into two piles—one that we are going to publish and the other that we are going to return—I wonder whether it would make any real difference to the journal or its readers if I exchanged one pile for another." (Fox 1965). However, evidence-based investigations reveal a significant effect of these decisions for a number of journals: (Lutz Bornmann and Daniel 2008) tracked 878 articles accepted by *Angewandte Chemie – International Edition* and 959 rejected articles that were then published elsewhere. Their results show that being accepted by *Angewandte Chemie – International Edition* increased the expected number of citations by up to 50 per cent. Similar results were obtained in investigations on the *American Journal of Neuroradiology* (McDonald, Cloft, and Kallmes 2007, 2009), *Cardiovascular Research* (Opthof et al. 2000), *F1000* (Lutz Bornmann and Leydesdorff 2013), and *Journal of Clinical Investigation* (Wilson 1978). Broader investigations were conducted by (Cicchetti 1991) and (Benda and Engels 2011), confirming these findings. Thus, citation analysis suggests that peer review tends to select 'the better pile' of manuscripts, which is subsequently cited more often.

#### 4.3 Grant decisions

Some studies have analysed grant peer reviews with the aim of assessing whether applicants who have been awarded funding were cited more often than unfunded applicants: van Leeuwen and Moed (2012) analysed the correlation between funding from three funding councils in the Netherlands and citation impact in the physical sciences, chemistry, and the geosciences. Successful applicants tended to generate a higher citation impact on the international level than those whose applications were rejected. A number of other studies (Armstrong et al. 1997; Lutz Bornmann and Daniel 2006, 2008; Cabezas-Clavijo et al. 2013) also revealed a positive correlation between grant peer review and citation impact. Other studies, however, showed no or only a low correlation between successful grant applications (DFG's Emmy Noether Programme (Germany), the Swedish Foundation for Strategic Research, and the Council for Social Scientific Research of the Netherlands) and subsequent citation impact (Hornbostel et al. 2009; Melin and Danell 2006; Van den Besselaar and Leydesdorff 2007). Thus, the results of the comparison of grant peer review and bibliometric indicators are mixed. Further to this, whether the selection of applicants might be a self-fulfilling prophecy was also questioned: "They reward prior excellence [...], but they also afford the successful

applicants resources that might enable them to do excellent scientific work in their future careers. If a study establishes significant performance advantages for accepted applicants but not rejected ones, one can argue [...] that the funding organization gives the fellows such an advantage in training, prestige, self-confidence, and so on that they later become superior scientists because of the fellowship, not because they were particularly promising at the point of application. Rather than picking the 'best' scientists, the selection committee might, in this view, create them." (L. Bornmann 2011)

#### 4.4 Tenure decision

Vieira, Cabral, and Gomes (2014) analysed 27 professor recruitment processes that took place between 2007 and 2011 at six Portuguese universities with 174 candidates who had published a total of 7,654 documents indexed in the Web of Science in the 10 years prior to their applications. The disciplines of chemistry, physics, biology, mathematics, mechanics, geology, and computer science were involved. When any two applicants in a given recruitment process were compared using a combination of two bibliometric indicators (an h-index variant and the percentage of highly cited documents), the outcome of the ranking of those two applicants by peers could be predicted in 75 per cent of cases. Jensen, Rouquier, and Croissant (2009) explored the correlation between bibliometric indicators (h index, h index divided by 'scientific age', number of citations, number of publications, and average number of citations per publication) and the results of a peer review process concerning the promotion of about 600 researchers at France's Centre Nationale de la Recherche Scientifique (CNRS). The authors found that "no single indicator is the best predictor for all disciplines. Overall, however, the Hirsch index h provides the least bad correlations, followed by the number of papers published. It is important to realize however, that even h is able to recover only half of the actual promotions. The number of citations or the mean number of citations per paper are definitely not good predictors of promotion." (Jensen, Rouquier, and Croissant 2009).

#### 4.5 Evaluation of research groups

There are several studies comparing peer review of research groups with bibliometric indicators. In general, these studies found a (sometimes weak) positive correlation. Wouters et al. (2015) explain "the imperfect correlations between bibliometric indicators and peer review (partly) by variation in qualitative peer-based judgements". For example, Aksnes and Taxt (2004) compared the peer ratings of 34 research groups at the University of Bergen (Norway) with a set of five bibliometric indicators. The highest Pearson's correlation was observed between peer ratings and an indicator called 'relative publication strategy'. It compares the average citation rate of the journals in which the group's articles were published with the average citation rates of the subfields covered by each journal. Meho and Sonnenwald (2000) analysed the relationship between citation ranking and peer evaluation in assessing senior Kurdologists' research performance. Normalized citation ranking and citation content analysis were highly correlated with peer ranking both for high-ranked and low-ranked senior scholars. Anthony van Raan et al. performed a number of investigations in this area: Nederhof and van Raan (1993) analysed the relationship between bibliometric indicators and peer review for six research groups in economics. Peer review and bibliometric findings were generally in agreement. Rinia et al. (1998) showed the correlation between different bibliometric indicators and the outcomes of peer review made by expert panels of physicists in the Netherlands. In the field of physics, they assessed a set of 56 research programmes with approx. 5,000 publications and 50,000 citations. They found the strongest correlation to be between bibliometric indicators and the judgement of the researchers and the research team. Later, van Raan investigated the correlation between standard bibliometric indicators and peer judgment for 147 chemistry research groups in the Netherlands (Van Raan 2006). He found that both h index and CPP/FCSm discriminate very well

between the sets of documents that received a rating of 3 ('satisfactory') and the sets of documents that received ratings of 4 ('good') and 5 ('excellent').

#### 4.6 National research assessments

A comparative analysis of citation indicators and peer ratings in the Italian research evaluation assessment Valutazione Triennale della Ricerca (VTR) revealed a significant correlation between the two approaches (Ancaiani et al. 2015), though this conclusion was later questioned for methodological reasons (Baccini and De Nicolao 2016). The latter study concluded that bibliometrics and peer review do not produce similar results. Both studies by Abramo, D'Angelo, and Caprasecca (2009) and Franceschet and Costantini (2011) describe positive correlation between peer decisions and the impact factor of the journals in which the documents were published at the Italian VTR, though they did identify differences between the scientific fields. A number of investigations have been performed with data from the British Research Assessment Exercise (RAE) and its successor, the Research Excellence Framework (REF). For the 1992 RAE, Oppenheim (1997) found for three subject areas (anatomy, archaeology, and genetics) statistically significant correlations between the total number of citations received, or the average number of citations per member of staff, and the RAE score. Similar results were obtained from the 2001 RAE data for the field of archaeology by (Norris and Oppenheim 2003). While some experts criticized the conclusions (e.g. (Warner 2000)), others replicated the results for the areas of political science and chemistry. However, "no single model will apply across science and non-science disciplines. Any metrics approach to performance evaluation has to use a discipline-specific suite of indicators." (Butler and Mcallister 2011). Butler and Mcallister (2011) used the mean citation rate, the department size, the research culture, and the presence of staff from the department being evaluated on the RAE panel as predictors of the peer decisions. They concluded that citations alone are no surrogate for a peer review. Work reviewing the bibliometrics of the RAE 2001 data came to the same conclusion. There were huge differences between the scientific fields, and even within a field with a statistically significant correlation between citation and RAE score such as chemistry, the ranking of institutions that emerged from the two measures was far from identical (Mahdi, D'Este, and Neely 2008). The subsequent RAE and REF assessment exercises have been subject of further investigations; for an overview, see Traag and Waltman (2019). For example, the outcome of REF 2014 was compared with fifteen bibliometric and altmetric indicators (HEFCE 2015). The authors found that the quantitative indicators with significant impact in predicting peer evaluation differed between scientific disciplines. Consistently high correlations were found for several metrics in clinical medicine, economics, and econometrics. The study concludes that metrics cannot provide a like-for-like replacement of REF peer review. However, the study does not analyse department-level average scores, which, according to (Traag and Waltman 2019), are more relevant for the REF.

## 5 Conclusion

Overall, most of the comparative studies found a moderately positive correspondence between peer review and bibliometric indicators, but the correlations identified have been far from perfect and have varied among the studies. Inter alia, the correlations depend on the scientific field, the bibliometric indicators, and the subject of the review.

- The results of studies focusing on grant decisions are mixed. While a number of studies revealed a positive correlation between grant peer review and citation impact, other studies showed no or only a low correlation between the success in grant applications and subsequent citation impact.
- The results of investigations on tenure decisions are not convincing either. At best, bibliometric indicators can predict the correct ranking of any two applicants in 75 per cent of



cases, which is only halfway between the actual result (100 per cent) and a random decision (50 per cent). Reviews of research group assessments generally revealed better correspondence between peer review and bibliometrics, often depending on the scientific field and the indicator in question.

- Investigations on national research assessments revealed results similar to the studies on research group assessments. As a rule of thumb, correlations in the area of science, technology, and medicine (STM) are better than in the social sciences and humanities, and correlations are better for field-normalized indicators than for basic indicators like the citation count.

Therefore, there is generally little empirical support for the hypothesis that bibliometrics reflects the same aspects of impact or research quality as peer review. However, the extent to which the correlation between the two approaches is considered sufficient depends on the nature and the goals of the evaluation. The statement by Abramo and D'Angelo (2011) regarding national research assessments could hold true for the evaluation of research groups as well: "Accepting that there is no one infallible evaluation method, the position of the authors is that for the natural and formal sciences, the bibliometric methodology is by far preferable to informed peer review." This may first appear as a daring thesis, but it can be justified for the following reasons:

- Peer review is far from perfect, as was shown, for example, in the section "Peer review vs. journal impact factor". Therefore, a deviation in the results of a bibliometric approach from a peer decision does not necessarily indicate that the bibliometric approach led to a 'wrong' result.
- Peer review is very labour intensive and hence expensive. This can be justified for tenure decisions, for example, but must be questioned for the production of rankings.
- While there are good reasons to be hesitant about utilizing bibliometric approaches as the only method for evaluating single persons (e.g. tenure, grant), the situation for assessing groups is different, and here bibliometric analyses could replace peer review.

## 6 Bibliography

- Abramo, G., and C. A. D'Angelo. 2011. "Evaluating research: from informed peer review to bibliometrics." *Scientometrics* 87 (3): 499-514. <https://doi.org/10.1007/s11192-011-0352-7>.
- Abramo, G., Ciriaco Andrea D'Angelo, and Alessandro Caprasecca. 2009. "Allocative efficiency in public research funding: Can bibliometrics help?" *Research Policy* 38 (1): 206-215. <https://doi.org/10.1016/j.respol.2008.11.001>.
- Aksnes, D. W., L. Langfeldt, and P. Wouters. 2019. "Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories." *Sage Open* 9 (1): 17. <https://doi.org/10.1177/2158244019829575>.
- Aksnes, D. W., and R. E. Taxt. 2004. "Peer reviews and bibliometric indicators: a comparative study at a Norwegian university." *Research Evaluation* 13 (1): 33-41. <https://doi.org/10.3152/147154404781776563>.
- Ancaiani, Alessio, Alberto F. Anfossi, Anna Barbara, Sergio Benedetto, Brigida Blasi, Valentina Carletti, Tindaro Cicero, Alberto Ciolfi, Filippo Costa, Giovanna Colizza, Marco Costantini, Fabio di Cristina, Antonio Ferrara, Rosa M. Lacatena, Marco Malgarini, Irene Mazzotta, Carmela A. Nappi, Sandra Romagnosi, and Serena Sileoni. 2015. "Evaluating scientific research in Italy: The 2004–10 research evaluation exercise." *Research Evaluation* 24 (3): 242-255. <https://doi.org/10.1093/reseval/rvv008>.
- Anonymous. 1989. "PEERS REVIEWED." *The Lancet* 333 (8647): 1115-1116. [https://doi.org/10.1016/S0140-6736\(89\)92390-8](https://doi.org/10.1016/S0140-6736(89)92390-8).
- Armstrong, P. W., M. M. Caverson, L. Adams, M. Taylor, and P. M. Olley. 1997. "Evaluation of the Heart and Stroke Foundation of Canada Research Scholarship Program: research productivity

- and impact." *The Canadian journal of cardiology* 13 (5): 507-516.  
<http://europepmc.org/abstract/MED/9179090>.
- Baccini, A., and G. De Nicolao. 2016. "Do they agree? Bibliometric evaluation versus informed peer review in the Italian research assessment exercise." *Scientometrics* 108 (3): 1651-1671.  
<https://doi.org/10.1007/s11192-016-1929-y>.
- Baldwin, Melinda. 2015. "Credibility, peer review, and Nature, 1945 -1990." *Notes and Records: the Royal Society Journal of the History of Science* 69 (3): 337-352.  
<https://doi.org/doi:10.1098/rsnr.2015.0029>.
- . 2018. "Scientific Autonomy, Public Accountability, and the Rise of "Peer Review" in the Cold War United States." *Isis* 109 (3): 538-558. <https://doi.org/10.1086/700070>.
- Benda, Wim G. G., and Tim C. E. Engels. 2011. "The predictive validity of peer review: A selective review of the judgmental forecasting qualities of peers, and implications for innovation in science." *International Journal of Forecasting* 27 (1): 166-182.  
<https://doi.org/10.1016/j.ijforecast.2010.03.003>.
- Bornmann, L. 2011. "Scientific Peer Review." *Annual Review of Information Science and Technology* 45: 199-245. <https://doi.org/10.1002/aris.2011.1440450112>.
- Bornmann, Lutz, and Hans-Dieter Daniel. 2006. "Selecting scientific excellence through committee peer review - A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants." *Scientometrics* 68 (3): 427-440.  
<https://doi.org/10.1007/s11192-006-0121-1>.
- . 2008. "Selecting manuscripts for a high-impact journal through peer review: A citation analysis of communications that were accepted by Angewandte Chemie International Edition, or rejected but published elsewhere." *Journal of the American Society for Information Science and Technology* 59 (11): 1841-1852. <https://doi.org/10.1002/asi.20901>.
- Bornmann, Lutz, and Loet Leydesdorff. 2013. "The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000." *Journal of Informetrics* 7 (2): 286-291. <https://doi.org/10.1016/j.joi.2012.12.003>.
- Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2010. "A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and Its Determinants." *PLOS ONE* 5 (12): e14331. <https://doi.org/10.1371/journal.pone.0014331>.
- Butler, Linda, and Ian Mcallister. 2011. "Evaluating University Research Performance Using Metrics." *European Political Science* 10 (1): 44-58. <https://doi.org/10.1057/eps.2010.13>.
- Cabezas-Clavijo, A., N. Robinson-Garcia, M. Escabias, and E. Jimenez-Contreras. 2013. "Reviewers' Ratings and Bibliometric Indicators: Hand in Hand When Assessing Over Research Proposals?" *Plos One* 8 (6): 12. <https://doi.org/10.1371/journal.pone.0068258>.
- Chism, Nancy Van Note, and Grady W. Chism. 2007. *Peer review of teaching : a sourcebook*. 2nd ed. Bolton, MA: Anker Pub. Co.
- Chop, Ines, and Maria Eberlein-Gonska. 2012. "Übersichtsartikel zum Peer Review Verfahren und seine Einordnung in der Medizin." *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 106 (8): 547-552. <https://doi.org/10.1016/j.zefq.2012.08.017>.
- Cicchetti, Domenic V. 1991. "The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation." *Behavioral and Brain Sciences* 14 (1): 119-135.  
<https://doi.org/10.1017/S0140525X00065675>.
- Cole, S, JR, and GA Simon. 1981. "Chance and consensus in peer review." *Science* 214 (4523): 881-886. <https://doi.org/10.1126/science.7302566>.
- de Knecht, Sicco. 2019. "So what about editor compensation?". Accessed 11.04.2019.  
<https://medium.com/@SiccodeKnecht/so-what-about-editor-compensation-89fccaf10d71>.
- Ertel, Paul P., and Melvin Gene Aldridge. 1977. *Medical peer review : theory and practice*. Saint Louis: Mosby.
- Eysenck, H. J., and S. B. G. Eysenck. 1992. "Peer review: Advice to referees and contributors." *Personality and Individual Differences* 13 (4): 393-399. [https://doi.org/10.1016/0191-8869\(92\)90066-X](https://doi.org/10.1016/0191-8869(92)90066-X).



- Fox, Theodore Fortescue. 1965. *Crisis in communication: the functions and future of medical journals*. London: Athlone Press.
- Franceschet, Massimo, and Antonio Costantini. 2011. "The first Italian research assessment exercise: A bibliometric perspective." *Journal of Informetrics* 5 (2): 275-291. <https://doi.org/10.1016/j.joi.2010.12.002>.
- Gallo, Stephen A., and Scott R. Glisson. 2018. "External Tests of Peer Review Validity Via Impact Measures." *Frontiers in Research Metrics and Analytics* 3 (22). <https://doi.org/10.3389/frma.2018.00022>.
- Gemma, Derrick. 2017. *The evaluators' eye : impact assessment and academic peer review*. 1st edition. ed. New York, NY: Springer Berlin Heidelberg.
- Godlee, Fiona. 2002. "Making Reviewers Visible: Openness, Accountability, and Credit." 287 (21): 2762-2765. <https://doi.org/10.1001/jama.287.21.2762>.
- Goldfinch, Shaun, and Kiyoshi Yamamoto. 2012. *Prometheus assessed? : Research measurement, peer review, and citation analysis*. Oxford: Chandos Publishing.
- Gould, Thomas H. P. 2013. *Do we still need peer review? : an argument for change*. Lanham: The Scarecrow Press, Inc.
- Hadian, Mehrnaz, Robert S. Iwrey, Conrad Meyer, and Patrick D. Souter. 2018. *What is . . . medical staff peer review?* Chicago, Illinois: American Bar Association.
- HEFCE. 2015. *The Metric Tide: Correlation analysis of REF2014 scores and metrics (Supplementary Report II to the Independent Review of the Role of Metrics in Research Assessment and Management)*.
- Helmholtz. 2018. *Annual Report*. [https://www.helmholtz.de/fileadmin/user\\_upload/04\\_mediathek/18\\_Helmholtz\\_Geschaeftebericht\\_ENGLISCH\\_epaper.pdf](https://www.helmholtz.de/fileadmin/user_upload/04_mediathek/18_Helmholtz_Geschaeftebericht_ENGLISCH_epaper.pdf).
- Hornbostel, Stefan, Susan Böhmer, Bernd Klingsporn, Jörg Neufeld, and Markus von Ins. 2009. "Funding of young scientist and scientific excellence." *Scientometrics* 79 (1): 171-190. <https://doi.org/10.1007/s11192-009-0411-5>.
- Jefferson, Tom, Elizabeth Wager, and Frank Davidoff. 2002. "Measuring the Quality of Editorial Peer Review." 287 (21): 2786-2790. <https://doi.org/10.1001/jama.287.21.2786>.
- Jennings, Charles G. 2006. "Quality and value: The true purpose of peer review." Springer Nature. Accessed 14.04.2019.
- Jensen, Pablo, Jean-Baptiste Rouquier, and Yves Croissant. 2009. "Testing bibliometric indicators by their prediction of scientists promotions." *Scientometrics* 78 (3): 467-479. <https://doi.org/10.1007/s11192-007-2014-3>.
- Lang, Daniel A. 1999. *Medical staff peer review : motivation and performance in the era of managed care*. Rev. ed. Chicago, Ill.: Health Forum, Inc.
- Mahdi, Surya, Pablo D'Este, and Andy Neely. 2008. *Citation counts: are they good predictors of RAE scores? A bibliometric analysis of RAE 2001*. (AIM Research). <https://www.bl.uk/collection-items/citation-counts-are-they-good-predictors-of-rae-scores-a-bibliometric-analysis-of-rae-2001>.
- McDonald, R.J., H.J. Cloft, and D.F. Kallmes. 2007. "Fate of Submitted Manuscripts Rejected from the American Journal of Neuroradiology: Outcomes and Commentary." *American Journal of Neuroradiology* 28 (8): 1430-1434. <https://doi.org/10.3174/ajnr.A0766>.
- . 2009. "Fate of Manuscripts Previously Rejected by the American Journal of Neuroradiology: A Follow-Up Analysis." *American Journal of Neuroradiology* 30 (2): 253-256. <https://doi.org/10.3174/ajnr.A1366>.
- McKiernan, Erin C., Lesley A. Schimanski, Carol Muñoz Nieves, Lisa Matthias, Meredith T. Niles, and Juan Pablo Alperin. 2019. "Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations." *PeerJ Preprints* 7: e27638v2. <https://doi.org/10.7287/peerj.preprints.27638v2>.
- Meho, Lokman I., and Diane H. Sonnenwald. 2000. "Citation ranking versus peer evaluation of senior faculty research performance: A case study of Kurdish scholarship." *Journal of the American*

- Society for Information Science* 51 (2): 123-138. [https://doi.org/10.1002/\(sici\)1097-4571\(2000\)51:2<123::Aid-asi4>3.0.Co;2-n](https://doi.org/10.1002/(sici)1097-4571(2000)51:2<123::Aid-asi4>3.0.Co;2-n).
- Melin, Göran, and Rickard Danell. 2006. "The top eight percent: Development of approved and rejected applicants for a prestigious grant in Sweden." *Science and Public Policy* 33 (10): 702-712. <https://doi.org/10.3152/147154306781778579>.
- Nederhof, A. J., and A. F. J. van Raan. 1993. "A bibliometric analysis of six economics research groups: A comparison with peer review." *Research Policy* 22 (4): 353-368. [https://doi.org/10.1016/0048-7333\(93\)90005-3](https://doi.org/10.1016/0048-7333(93)90005-3).
- Norris, Michael, and Charles Oppenheim. 2003. "Citation counts and the Research Assessment Exercise V: Archaeology and the 2001 RAE." *Journal of Documentation* 59 (6): 709-730. <https://doi.org/10.1108/00220410310698734>.
- Oppenheim, Charles. 1997. "The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology." *Journal of Documentation* 53 (5): 477-487. <https://doi.org/doi:10.1108/EUM0000000007207>.
- Opthof, Tobias, Fred Furstner, Michael van Geer, and Ruben Coronel. 2000. "Regrets or no regrets? No regrets! The fate of rejected manuscripts." *Cardiovascular Research* 45 (1): 255-258. [https://doi.org/10.1016/S0008-6363\(99\)00339-9](https://doi.org/10.1016/S0008-6363(99)00339-9).
- Paltridge, Brian. 2017. *The discourse of peer review : reviewing submissions to academic journals*. London, United Kingdom: Palgrave Macmillan.
- Pierie, Jean-Pierre E. N., Henk C. Walvoort, and A. John P. M. Overbeke. 1996. "Readers' evaluation of effect of peer review and editing on quality of articles in the Netherlands Tijdschrift voor Geneeskunde." *The Lancet* 348 (9040): 1480-1483. [https://doi.org/10.1016/S0140-6736\(96\)05016-7](https://doi.org/10.1016/S0140-6736(96)05016-7).
- Retraction Watch. 2016. "Dear peer reviewer, you stole my paper: An author's worst nightmare." Accessed 12.04.2019. <https://retractionwatch.com/2016/12/12/dear-peer-reviewer-stole-paper-authors-worst-nightmare/>.
- Rinia, E. J., Th N. van Leeuwen, H. G. van Vuren, and A. F. J. van Raan. 1998. "Comparative analysis of a set of bibliometric indicators and central peer review criteria: Evaluation of condensed matter physics in the Netherlands." *Research Policy* 27 (1): 95-107. [https://doi.org/https://doi.org/10.1016/S0048-7333\(98\)00026-2](https://doi.org/https://doi.org/10.1016/S0048-7333(98)00026-2).
- Roy, Rustum. 1985. "Funding science: The real defects of peer-review and an alternative to it." *Science, Technology, & Human Values* 52 (3): 73-81. <https://doi.org/10.1177/016224398501000309>.
- Samson, Sue, and Donna E. McCrea. 2008. "Using peer review to foster good teaching." *Reference Services Review* 36 (1): 61-70. <https://doi.org/10.1108/00907320810852032>.
- Schultz, David M. 2010. "Rejection Rates for Journals Publishing in the Atmospheric Sciences." *Bulletin of the American Meteorological Society* 91 (2): 231-244. <https://doi.org/10.1175/2009bams2908.1>.
- Smith, Richard. 2006. "Peer review: a flawed process at the heart of science and journals." *Journal of the Royal Society of Medicine* 99 (4): 178-182. <https://doi.org/10.1258/jrsm.99.4.178>.
- Snaveley, Loanne, and Nancy Dewald. 2011. "Developing and Implementing Peer Review of Academic Librarians' Teaching: An Overview and Case Report." *Journal of Academic Librarianship* 37 (4): 343-351. <https://doi.org/10.1016/j.acalib.2011.04.009>.
- Traag, V. A., and L. Waltman. 2019. "Systematic analysis of agreement between metrics and peer review in the UK REF." *Palgrave Communications* 5 (1): 29. <https://doi.org/10.1057/s41599-019-0233-x>.
- Van den Besselaar, Peter, and Loet Leydesdorff. 2007. *Past performance as predictor of successful grant applications: A case study. Report to the board of the Netherlands Social Science Research council (MaGW/NWO)*.
- van Leeuwen, Thed N., and Henk F. Moed. 2012. "Funding decisions, peer review, and scientific excellence in physical sciences, chemistry, and geosciences." *Research Evaluation* 21 (3): 189-198. <https://doi.org/10.1093/reseval/rvs009>.

- Van Raan, A. F. J. 2006. "Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups." *Scientometrics* 67 (3): 491-502.  
<https://doi.org/10.1556/Scient.67.2006.3.10>.
- Van Valey, Thomas L. 2011. *Peer review of teaching : lessons from and for departments of sociology*. Washington, DC: American Sociological Association.
- Vieira, Elizabeth S., José A.S. Cabral, and José A.N.F. Gomes. 2014. "Definition of a model based on bibliometric indicators for assessing applicants to academic positions." *Journal of the Association for Information Science and Technology* 65 (3): 560-577.  
<https://doi.org/10.1002/asi.22981>.
- Vieira, Elizabeth S., and José A.N.F. Gomes. 2018. "The peer-review process: The most valued dimensions according to the researcher's scientific career." *Research Evaluation* 27 (3): 246-261. <https://doi.org/10.1093/reseval/rvy009>.
- Walsh, Elizabeth, Maeve Rooney, Louis Appleby, and Greg Wilkinson. 2000. "Open peer review: A randomised controlled trial." *British Journal of Psychiatry* 176 (1): 47-51.  
<https://doi.org/10.1192/bjp.176.1.47>.
- Warner, Julian. 2000. "A Critical Review of the Application of Citation Studies to the Research Assessment Exercises." *Journal of Information Science* 26: 453-460.  
<https://doi.org/10.1177/0165555150002600607>.
- Weller, Ann C. 2001. *Editorial peer review : its strengths and weaknesses*. ASIST monograph series. Medford, N.J.: Information Today.
- Wilson, J. D. 1978. "Peer review and publication. Presidential address before the 70th annual meeting of the American Society for Clinical Investigation, San Francisco, California, 30 April 1978." *The Journal of Clinical Investigation* 61 (6): 1697-1701.  
<https://doi.org/10.1172/JCI109091>.
- Wouters, Paul, Mike Thelwall, Kayvan Kousha, Ludo Waltman, Sarah de Rijcke, Alex Rushforth, and Thomas Franssen. 2015. *The Metric Tide: Literature Review (Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management)*. (HEFCE).
- Zuckerman, Harriet, and Robert K. Merton. 1971. "Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system." *Minerva* 9 (1): 66-100.  
<https://doi.org/10.1007/bf01553188>.