

Chapter 8

Peer Review and Bibliometric: Potentials and Problems

Lutz Bornmann

8.1 Introduction

Among the various human activities, activities in science are those that are the most subject to evaluation by peers (Laloë and Mosseri 2009). Such evaluations determine, among ranking positions of universities, who gets which job, who gets tenure, and who gets which awards and honors (Feist 2006). For the THE – QS World University Rankings, the assessment by peers is the centerpiece of the ranking process; peer review is also a major indicator in the US News & World Report rankings (Enserink 2007). “By defining losers and winners in the competition for positions, grants, publication of results, and all kinds of awards, peer review is a central social control institution in the research community” (Langfeldt 2006: 32). Research evaluation systems in the various countries of the world (e.g., the British research assessment exercise) are normally based on peer review. The edited book of Whitley and Gläser (2007) shows how these systems are changing the organization of scientific knowledge production and universities in the countries involved (Moed 2008).

Aside from the selection of manuscripts for publication in journals, the most common contemporary application of peer review in scientific research is for the selection of fellowship and grant applications. Peers or colleagues, asked to evaluate applications or manuscripts in a peer review process, take on the responsibility for assuring high standards in various research disciplines. Although peers active in the same field might be blind-sided by adherence to the same specialist group, they “are said to be in the best position to know whether quality standards have been met and a contribution to knowledge made” (Eisenhart 2002: 241). Peer evaluation in

L. Bornmann (✉)
Max Planck Society, Hofgartenstr. 8, D-80539 Munich, Germany
e-mail: bornmann@gv.mpg.de

research thus entails a process by which a selective jury of equals, active in a given scientific field, convenes to evaluate the undertaking of scientific activity or its outcomes. Such a jury of equals may be consulted as a group or individually, without the need for personal contacts among the evaluators. The peer review process lets the active producers of science, the experts, become the “gatekeepers” of science (McClellan 2003).

Proponents of the peer review system argue that it is more effective than any other known instrument for self-regulation in science. Putting it into a wider context, according to the critical rationalism of Popper (1961) intellectual life and institutions should be arranged to provide “maximum criticism, in order to counteract and eliminate as much intellectual error as possible” (Bartley 1984: 113). Evidence supports the view that peer review improves the quality of the reporting of research results (Goodman et al. 1994; Pierie et al. 1996). As a proponent of peer review, Abelson writes (1980): “The most important and effective mechanism for attaining good standards of quality in journals is the peer review system” (p. 62). According to Shatz (2004) journal peer review “motivates scholars to produce their best, provides feedback that substantially improves work which is submitted, and enables scholars to identify products they will find worth reading” (p. 30).

Critics of peer review argue that (1) reviewers rarely agree on whether or not to recommend that a manuscript be published or a research grant be awarded, thus making for poor *reliability* of the peer review process; (2) reviewers’ recommendations are frequently biased, that is, judgments are not based solely on scientific merit, but are also influenced by personal attributes of the authors, applicants, or the reviewers themselves (where the *fairness* of the process is not given); and (3) the process lacks *predictive validity*, since there is little or no relationship between the reviewers’ judgments and the subsequent usefulness of the work to the scientific community, as indicated by the frequency of citations of the work in later scientific papers. According to Butler (2007), the assessment by peers as an indicator in the US News & World Report university ranking implies a false precision and authority. For further criticisms on scientific peer review see Hames (2007) and Schmelkin (2006).

In recent years, a number of published studies have addressed these criticisms raised about scientific peer review. From the beginning, this research on peer review has focused on the evaluation of manuscripts and (fellowship or grant) applications.

“The peer review process that scholarly publications undergo may be interpreted as a sign of ‘quality.’ But to many, a publication constitutes nothing more than an ‘offer’ to the scientific community. It is the subsequent reception of that offer that certifies the actual ‘impact’ of a publication” (Schneider 2009: 366). Formal citations are meant to show that a publication has made use of the contents of other publications (research results, others’ ideas, and so on). Citation counts (the number of citations) are used in research evaluation as an indicator of the impact of the research: “The *impact* of a piece of research is the degree to which it has been useful to other researchers” (Shadbolt et al. 2006: 202). According to the Research Evaluation and Policy Project (2005), there is an emerging trend to regard impact, the measurable part of quality, as a proxy measure for quality in total. For Lindsey,

citations are “our most *reliable* convenient measure of quality in science – a measure that will continue to be widely used” (Lindsey 1989: 201).

In research evaluation, citation analyses have been conducted for assessment of national science policies and disciplinary development (e.g., Lewison 1998; Oppenheim 1995, 1997; Tijssen et al. 2002), departments and research laboratories (e.g., Bayer and Folger 1966; Narin 1976), books and journals (e.g., Garfield 1972; Nicolaisen 2002), and individual scientists (e.g., Cole and Cole 1973; Garfield 1970). Besides peer review with a 40% weighting, the THE – QS World University Rankings gives the indicator “citations per faculty” a 20% weighting. The Leiden Ranking system is entirely based on bibliometric indicators (Enserink 2007).

Citation counts are attractive raw data for the evaluation of research output. Because they are “unobtrusive measures that do not require the cooperation of a respondent and do not themselves contaminate the response (i.e., they are non-reactive)” (Smith 1981: 84), citation rates are seen as an *objective* quantitative indicator for scientific success and are held to be a valuable complement to qualitative methods for research evaluation, such as peer review (Daniel 2005; Garfield and Welljamsdorff 1992). Scientific “reward came primarily in the form of recognition rather than money, an insight that helps account for the importance scientists place upon citation as a reward system ... This idea of citation as a kind of stand-in for direct economic reward – what is sometimes called the citation credit cycle – is often seen as a feature of academic reward generally” (Kellogg 2006: 3).

However, out in the early 1970s, Eugene Garfield, the founder of the Institute of Scientific Information (ISI, now Thomson Reuters, Philadelphia, PA, USA) pointed out that citation counts are a function of many variables besides scientific quality (Garfield 1972). In a recently published paper, Laloë and Mosseri (2009) state that bibliometric methods “do contain information about scientific quality, but this ‘signal’ is buried in a ‘noise’ created by a dependence on many other variables” (p. 27). Up to now, a number of variables that generally influence citation counts have emerged in bibliometric studies. Lawani (1986) and other researchers established, for example, that there is a positive relation between the number of co-authors of a publication and its citation counts; a higher number of co-authors is usually associated with a higher number of citations. Based on the findings of these studies, the number of co-authors and other general influencing factors should be taken into consideration in evaluative bibliometric studies.

Since research evaluation is an area of increasing importance, it is necessary that the application of peer review and impact measures (citation counts) is done well and professionally (see here de Vries et al. 2009). For that, background information about empirical findings on both evaluation instruments is necessary (especially findings that are related to their problems). In Sect. 8.2 of this chapter, an overview is provided on studies that have conducted meta-evaluations of peer review procedures, because a literature search found no empirical studies on peer review in the context of university rankings, Sect. 8.2 focuses on journal, fellowship, and grant peer review. In general, the results are applicable to the use of peer review in the context of university rankings. Sect. 8.3 gives an overview on studies that have investigated citation counts to identify general influencing factors.

8.2 Research on Journal, Fellowship, and Grant Peer Review

8.2.1 Agreement Among Reviewers (*Reliability*)

“In everyday life, intersubjectivity is equated with realism” (Ziman 2000: 106). The scientific discourse is also distinguished by a striving for consensus. Scientific activity would clearly be impossible unless scientists could come to similar conclusions. According to Wiley (2008) “just as results from lab experiments provide clues to an underlying biological process, reviewer comments are also clues to an underlying reality (they did not like your grant for some reason). For example, if all reviewers mention the same point, then it is a good bet that it is important and real.” An established consensus among scientists must of course be a voluntary one achieved under conditions of free and open criticism (Ziman 2000). The norms of the ethos of science make these conditions possible and regulate them (Merton 1942): The norms of communalism (scientific knowledge should be made public knowledge) and universalism (knowledge claims should be judged impersonally, independently of their source) envisage eventual agreement. “But the norm of ‘organized skepticism’, which energizes critical debates, rules out any official procedure for closing them. Consensus and dissensus are thus promoted simultaneously” (Ziman 2000: 255) by the norms of the ethos of science.

If a submission (manuscript or application) meets scientific standards and contributes to the advancement of science, one would expect that two or more reviewers will agree on its value. This, however, is frequently not the case. Ernst et al. (1993) offer a dramatic demonstration of the unreliability of the journal peer review process. Copies of one paper submitted to a medical journal were sent simultaneously to 45 experts. They were asked to express their opinion of the paper with the journal’s standard questionnaire judging eight quality criteria on a numerical scale from 5 (excellent) to 1 (unacceptable). The 31 correctly filled forms demonstrated poor reliability with extreme judgments ranging from “unacceptable” to “excellent” for most criteria. The results of studies on reliability in journal peer review indicate that the levels of inter-reviewer agreement, when corrected for chance, generally fall in the range from 0.20 to 0.40 (Bornmann 2011), which indicates a relatively low level of reviewer agreement.

Reviewer disagreement is not always seen as a negative factor however, as many see it as a positive method of evaluating a manuscript from a number of different perspectives. If reviewers are selected for their opposing viewpoints or expertise, a high degree of reviewer agreement should not be expected. It can even be argued that too much agreement is in fact a sign that the review process is not working well, that reviewers are not properly selected for diversity, and that some are redundant. Whether the comments of reviewers are in fact based on different perspectives is a question that has been examined by only a few empirical studies (Weller 2002). One study, for example, showed that reviewers of the same manuscript simply commented on different aspects of the manuscript: “In the typical case, two reviews of the same paper had no critical point in common ... [T]hey wrote about different

topics, each making points that were appropriate and accurate. As a consequence, their recommendations about editorial decisions showed hardly any agreement” (Fiske and Fogg 1990: 591).

The fate of a manuscript depends on which small sample of reviewers influences the editorial decision, as research such as that of Bornmann and Daniel (2009a, 2010) for the *Angewandte Chemie International Edition* (AC-IE) indicates. In AC-IE’s peer review process, a manuscript is generally published only if two reviewers rate the results of the study as important and also recommend publication in the journal (what the editors have called the “clear-cut” rule). Even though the “clear-cut” rule is based on two reviewer reports, submitted manuscripts generally go out to three reviewers in total. An editor explains this process in a letter to an author as follows: “Many papers are sent initially to three referees (as in this case), but in today’s increasingly busy climate there are many referees unable to review papers because of other commitments. On the other hand, we have a responsibility to authors to make a rapid and fair decision on the outcome of papers.” For 23% of those manuscripts, for which a third reviewer report arrived after the editorial decision was made (37 of 162), this rule would have led to a different decision if the third report had replaced either of the others. Consequently, even if the editor considered all three reviewers to be suitable to review a manuscript, the editor would have needed to make a different decision based on the changed situation.

8.2.2 *Fairness of the Peer Review Process*

According to Merton (1942) the functional goal of science is the expansion of potentially true and secure knowledge. To fulfill this function in society, the ethos of science was developed. The norm of universalism prescribes that the evaluation of scientific contributions should be based upon objective scientific criteria. Journal submissions or grant applications are not supposed to be judged according to the attributes of the author/applicant or the personal biases of the reviewer, editor, or program manager (Ziman 2000). “First, universalism requires that when a scientist offers a contribution to scientific knowledge, the community’s assessment of the validity of that claim should not be influenced by personal or social attributes of the scientist ... Second, universalism requires that a scientist be fairly rewarded for contributions to the body of scientific knowledge ... Particularism, in contrast, involves the use of functionally irrelevant characteristics, such as sex and race, as a basis for making claims and gaining rewards in science” (Long and Fox 1995: 46). To the degree that particularism influences how claims are made and rewards are gained, the fairness of the peer review process is at risk (Godlee and Dickersin 2003).

Ever since Kuhn (1962) discussed the significance of different scientific views or paradigmatic views for the evaluation of scientific contributions in his seminal work *The structure of scientific revolutions* (see here also Mallard et al. 2009), researchers have expressed increasing doubt about the norm-ruled objective evaluation of scientific work (Hemlin 1996). Above all, proponents of social

constructivism have expressed such doubts since the 1970s. For Cole (1992) the research of the constructivists supports a new view of science which casts doubt on the existence of a set of rational criteria. The most valuable of insights into scientists' actions, social constructivist research, according to Sismondo (1993), has brought about the recognition that "social objects in science exist and act as causes of, and constraints on, scientists' actions" (p. 548). Because reviewers are human, factors which cannot be predicted, controlled, or standardized influence their writing of reviews, according to Shashok (2005).

Reviews of peer review research (Hojat et al. 2003; Owen 1982; Pruthi et al. 1997; Ross 1980; Sharp 1990; Wood and Wessely 2003) name up to 25 potential sources of bias in peer review. In these studies, it is usual to call any feature of an assessor's cognitive or attitudinal mind-set that could interfere with an objective judgment, a bias (Shatz 2004). Factors that appear to bias assessors' objective judgments with respect to a manuscript or an application include nationality, gender of the author or applicant, and the area of research from which the work originates. Other studies show that replication studies and research that lead to statistically insignificant findings stand a rather low chance of being judged favorably by peer reviewers.

Research on bias in peer review faces two serious problems. First, the research findings on bias are inconsistent. For example, some studies investigating gender bias in journal review processes point out that women scientists are at a disadvantage. However, a similar number of studies report no gender effects or mixed results. Second, it is almost impossible to establish unambiguously whether work from a particular group of scientists (e.g., junior or senior scientists) receives better reviews and thus a higher acceptance rate due to preferential biases affecting the review and decision-making process, or if favorable review and favorable judgments in peer review are simply a consequence of the high scientific quality of the corresponding manuscripts or applications.

Presumably, it will never be possible to eliminate all doubts regarding the fairness of the review process. Because reviewers are human, their behavior – whether performing their salaried duties, enjoying their leisure time, or writing reviews – is influenced by factors that cannot be predicted, controlled or standardized (Shashok 2005). Therefore, it is important that the peer review process should be further studied. Any evidence of bias in judgments should be uncovered for purposes of correction and modification of the process (Geisler 2001; Godlee and Dickersin 2003).

8.2.3 Predictive Validity of the Peer Review Process

The goal for peer review of grant/ fellowship applications and manuscripts is usually to select the "best" from among the work submitted (Smith 2006). In investigating the predictive validity of the peer review process, the question arises as to whether this goal is actually achieved, that is, whether indeed the "best" applications or manuscripts are funded or published. The validity of judgments in peer review is often questioned. For example, the former editor of the journal *Lancet*,

Sir Theodore Fox (1965), writes on the validity 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” (p. 8). The selection function is considered to be a difficult research topic to investigate. According to Jayasinghe et al. (2001) and Figueredo (2006), there exists no mathematical formula or uniform definition as to what makes a manuscript “worthy of publication,” or what makes a research proposal “worthy of funding” (see also Smith 2006).

For the investigation of the predictive validity of the peer review process, the impact of papers accepted or rejected (but published elsewhere) in peer reviewed journals, or the impact of papers that were published by applicants whose proposals were either accepted or rejected in grant or fellowship peer reviews, are compared. Because the number of citations of a publication reflects its international impact (Borgman and Furner 2002; Nicolaisen 2007) and because of the lack of other operationalizable indicators, it is a common approach in peer review research to evaluate the success of the process on the basis of citation counts (see Sect. 8.3). Scientific judgments on submissions (manuscripts or applications) are said to show predictive validity in peer review research, if the citation counts of manuscripts accepted for publication (or manuscripts published by accepted applicants) and manuscripts rejected by a journal but then published elsewhere (or manuscripts published by rejected applicants) differ statistically significantly.

Up until now, only a few studies have conducted analyses which examine citation counts from individual papers as the basis for assessing predictive validity in peer reviews. A literature research found only six empirical studies on the level of predictive validity associated with the journal peer review process. Research in this area is extremely labor-intensive, since a validity test requires information and citation counts regarding the fate of rejected manuscripts (Bornstein 1991). The editor of the *Journal of Clinical Investigation* (Wilson 1978) has undertaken his own investigation into the question of predictive validity. Daniel (1993) and Bornmann and Daniel (2008a, b) investigated the peer review process of AC-IE, and Opthof et al. (2000) did the same for *Cardiovascular Research*. McDonald et al. (2009) and Bornmann et al. (2010) examined the predictive validity of the editorial decisions for the *American Journal of Neuroradiology* and *Atmospheric Chemistry and Physics*. All six studies confirmed that the editorial decisions (acceptance or rejection) for the various journals appear to reflect a rather high degree of predictive validity, if citation counts are employed as validity criteria.

Eight studies on the assessment of citation counts, as a basis of predictive validity in selection decisions in fellowship or grant peer reviews, have been published in recent years according to a literature search. The studies by Armstrong et al. (1997) on the Heart and Stroke Foundation of Canada (HSFC, Ottawa), the studies by Bornmann and Daniel (2005b, 2006) on the Boehringer Ingelheim Fonds (Heidesheim, Germany), and by Bornmann et al. (2008) on the European Molecular Biology Organization (Heidelberg, Germany), and the study of Reinhart (2009) on the Swiss National Science Foundation (Bern) confirm the predictive validity of the

selection decisions, whereas the studies by Hornbostel et al. (2009) on the Emmy Noether Programme of the German Research Foundation (Bonn) and by Melin and Danell (2006) on the Swedish Foundation for Strategic Research (Stockholm) showed no significant differences between the performance of accepted and rejected applicants. Van den Besselaar and Leydesdorff (2007) report on contradictory results regarding the Council for Social Scientific Research of the Netherlands Organization for Scientific Research (Den Haag). The study by Carter (1982) investigated the association between (1) assessments given by the reviewers for the National Institutes of Health (Bethesda, MD, USA) regarding applicants for research funding, and (2) the number of citations, which articles in journals produced under the grants have obtained. This study showed that better votes in fact correlate with more frequent citations; however, the correlation coefficient was low.

Unlike the clearer results for journal peer reviews, contradictory results emerge in research on fellowship or grant peer reviews. Some studies confirm the predictive validity of peer reviews, while the results of other studies leave room for doubt about their predictive validity.

8.3 Research on Citation Counts as Bibliometric Indicator

The research activity of a group of scientists, publication of their findings, and citation of the publications by colleagues in the field are all social activities. This means that citation counts for the group's publications are not only an indicator of the impact of their scientific work on the advancement of scientific knowledge (as stated by the *normative theory* of citing; see a description of the theories of citing in the next section). According to the *social constructivist view* on citing, citations also reflect (social) factors that do not have to do with the accepted conventions of scholarly publishing (Bornmann and Daniel 2008c). "There are 'imperfections' in the scientific communications system, the result of which is that the importance of a paper may not be identical with its impact. The 'impact' of a publication describes its actual influence on surrounding research activities at a given time. While this will depend partly on its importance, it may also be affected by such factors as the location of the author, and the prestige, language, and availability of the publishing journal" (Martin and Irvine 1983: 70). Bibliometric studies published in recent years have revealed the general influence of this and a number of other factors on citation counts (Peters and van Raan 1994).

8.3.1 *Theoretical Approaches to Explaining Citing*

Two competing theories of citing have been developed in past decades, both of them situated within broader social theories of science. One is often denoted as the normative theory of citing and the other as the social constructivist view of citing.

The *normative theory*, following Robert K. Merton's sociological theory of science (Merton 1973), basically states that scientists give credit to colleagues whose work they use by citing that work. Thus, citations represent intellectual or cognitive influence on scientific work. Merton (1988) expressed this aspect as follows: "The reference serves both instrumental and symbolic functions in the transmission and enlargement of knowledge. Instrumentally, it tells us of work we may not have known before, some of which may hold further interest for us; symbolically, it registers in the enduring archives the intellectual property of the acknowledged source by providing a pellet of peer recognition of the knowledge claim, accepted or expressly rejected, that was made in that source" (p. 622, see also Merton 1957; Merton 1968).

The *social constructivist view* on citing is grounded in the constructivist sociology of science (see, e.g., Collins 2004; Knorr-Cetina 1981; Latour and Woolgar 1979). This view casts doubt on the assumptions of normative theory and questions the validity of evaluative citation analysis. Constructivists argue that the cognitive content of articles has little influence on how they are received. Scientific knowledge is socially constructed through the manipulation of political and financial resources, and the use of rhetorical devices (Knorr-Cetina 1991). For this reason, citations cannot be satisfactorily described unidimensionally through the intellectual content of the article itself. The probability of being cited depends on many factors that are not related to the accepted conventions of scholarly publishing. In the next section, an overview of these factors is given.

8.3.2 *Factors that Influence Citation Counts in General*

8.3.2.1 **Time-Dependent Factors**

Due to the exponential increase in scientific output, citations become more probable from year to year. Beyond that, it has been shown that the more frequently a publication is cited, the more frequently it will be cited in future; in other words, the expected number of future citations is a linear function of the current number. Cozzens (1985) calls this phenomenon "success-breeds-success," and it holds true not only for highly-cited publications, but also for highly-cited scientists (Garfield 2002). However, according to Jensen et al. (2009) "the assumption of a constant citation rate *unlimited in time* is not supported by bibliometric data" (p. 474).

8.3.2.2 **Field-Dependent Factors**

Citation practices vary between science and social science fields (Castellano and Radicchi 2009; Hurt 1987; Radicchi et al. 2008) and even within different areas (or clusters) within a single subfield (Bornmann and Daniel 2009b). In some fields, researchers cite recent literature more frequently than in others. As the chance of

being cited is related to the number of publications in the field, small fields attract far fewer citations than more general fields (King 1987).

8.3.2.3 Journal-Dependent Factors

Ayres and Vars (2000) found that the first article in the journal tended to produce more citations than the later ones, perhaps because the editors recognized such articles to be especially important. Stewart (1983) argued that the citation of an article may depend on the frequency of publication of journals containing related articles. Furthermore, journal accessibility, visibility, and internationality as well as the impact, quality, or prestige of the journal may influence the probability of citations (Judge et al. 2007; Larivière and Gingras 2010; Leimu and Koricheva 2005).

8.3.2.4 Article-Dependent Factors

Citation characteristics of methodology articles, review articles, research articles, letters, and notes as well as articles, chapters, and books differ considerably (Lundberg 2007). There is also a positive correlation between the citation frequency of publications and (1) the number of co-authors of the work (Lansingh and Carter 2009), and (2) the number (Fok and Franses 2007) and the impact (Boyack and Klavans 2005) of the references within the work. Moreover, as longer articles have more content that can be cited than do shorter articles, the sheer size of an article influences whether it is cited (Hudson 2007).

8.3.2.5 Author- /Reader-Dependent Factors

The language a paper is written in (Kellsey and Knieval 2004; Lawani 1977) and cultural barriers (Carpenter and Narin 1981; Menou 1983) influence the probability of citations. Results from Mählck and Persson (2000), White (2001), and Sandström et al. (2005) show that citations are affected by social networks, and that authors cite primarily works by authors with whom they are personally acquainted. Cronin (2005) finds this hardly surprising, as it is to be expected that personal ties become manifest and strengthened, resulting in greater reciprocal exchange of citations over time.

8.3.2.6 Literature- and Citation Database-Dependent Factors

Free online availability of publications influences the probability of citations (Lawrence 2001; McDonald 2007). Citation analyses cannot be any more accurate than the raw material used (Smith 1981; van Raan 2005b). The incorrect citing of sources is unfortunately far from uncommon. Evans et al. (1990) checked the references in papers in three medical journals and determined that 48% were

incorrect: “The data support the hypothesis that authors do not check their references or may not even read them” (p. 1353). In a similar investigation, Eichorn and Yankauer (1987) found that “thirty-one percent of the 150 references had citation errors, one out of 10 being a major error (reference not locatable)” (p. 1011). Unver et al. (2009) found errors in references “in about 30% of current physical therapy and rehabilitation articles” (p. 744). Furthermore, the data in the literature data bases like Web of Science (WoS, Thomson Reuters) or Scopus (Elsevier) are not “homogeneous, since the entry of data has fluctuated in time with the persons in charge of it. It, therefore, requires a specialist to make the necessary series of corrections” (Laloë and Mosseri 2009: 28). Finally, according to Butler (2007) “Thomson Scientific’s [now Thomson Reuters] ISI citation data are notoriously poor for use in rankings; names of institutions are spelled differently from one article to the next, and university affiliations are sometimes omitted altogether. After cleaning up ISI data on all UK papers for such effects, the Leeds-based consultancy Evidence Ltd. found the true number of papers from the University of Oxford, for example, to be 40% higher than listed by ISI, says director Jonathan Adams” (p. 514, see also Bar-Ilan 2009). Errors in these data are especially serious, as most of the rankings are based on Thomson Reuter’s data (Buela-Casal et al. 2007).

8.4 Discussions

Buela-Casal et al. (2007) presented a comparative study of four well-known international university rankings. Their results show that generally peer review and citation counts play an important role as indicators in these rankings. Although university rankings are a growing phenomenon in higher education worldwide (Merisotis and Sadlak 2005), there is surprisingly little empirical research on the use of these dominating indicators. The research on peer review and citation counts (still) refers to other areas. However, as the results of this research are generalizable, this chapter has provided a research overview including the most important studies.

Against the backdrop of these studies, it can be assumed that peer assessments given for rankings are affected by disagreements among independent peers as well as biases and a lack of predictive validity: (1) One and the same university will be assessed differently by independent peers; (2) other criteria than scientific quality will influence the universities’ assessments; (3) the assessments might not be correlated with other indicators of scientific quality. Referring to citation counts, the research points out that this impact measure is affected by some general influencing factors. Thus, citation counts only measure an aspect of the scientific quality of universities. In the following paragraphs, we will summarize and discuss the most important findings presented in Sects. 8.2 and 8.3.

In recent years, a number of published studies have taken up and investigated the criticisms that have been raised against the scientific peer review process. Some important studies were presented in Sect. 8.2. To recapitulate the study results

published so far on the *reliability* of peer review: Most studies report a low level of agreement between reviewers' judgments. However, very few studies have investigated reviewer agreement with the purpose of identifying the actual reasons behind reviewer disagreement (e.g., by carrying out comparative content analyses of reviewers' comment sheets). LaFollette (1992), for example, noted the scarcity of research on such questions as how reviewers apply standards and the specific criteria established for making a decision. In-depth studies that address these issues might prove to be fruitful avenues for future investigation (Weller 2002). This research should primarily dedicate itself to the dislocational component in the judgment of reviewers as well as differences in strictness or leniency in reviewers' judgments (Eckes 2004; Lienert 1987).

Although reviewers like to believe that they choose the "best" based on objective criteria, "decisions are influenced by factors – including biases about race, sex, geographic location of a university, and age – that have nothing to do with the quality of the person or work being evaluated" (National Academy of Sciences 2006). Considering that peers are not prophets but ordinary human beings with their own opinions, strengths, and weaknesses (Ehse 2004), a number of studies have already worked on *potential sources of bias* in peer review. Although numerous studies have shown an association between potential sources of bias and judgments in peer review and thus called into question the fairness of the process itself, the research on these biases faces two fundamental problems that make generalization of the findings difficult. On the one hand, the various studies have yielded quite heterogeneous results. Some studies have proven the indisputable effects of potential sources of bias; in other studies, they showed moderate or slight effects. A second principal problem that affects bias research in general is the pervasive lack of experimental studies. This shortage makes it impossible to establish unambiguously whether work from a particular group of scientists receives better reviews due to biases in the review and decision-making process, or if favorable reviews and greater success in the selection process are simply a consequence of the scientific merit of the corresponding group of proposals or manuscripts.

The few studies, which have examined the *predictive validity* of journal peer review on the basis of citation counts, confirm that a peer review represents a quality filter and works as an instrument for the self-regulation of science. Concerning fellowship or grant peer reviews, there are more studies which have investigated the predictive validity of selection decisions on the basis of citation counts. Compared with journal peer reviews, these studies have provided heterogeneous results; some studies can confirm the predictive validity of peer reviews, whereas the results of other studies leave that in doubt.

The heterogeneous results on fellowship and grant peer review can be attributed to the fact that "funding decisions are inherently speculative because the work has not yet been done" (Stamps 1997: 4). Whereas in a journal peer review the *results* of the research are assessed, a grant and fellowship peer review is principally an evaluation of the *potential* of the proposed research (Bornmann and Daniel 2005a). Evaluating the application involves deciding whether the proposed research is significant, determining whether the specific plans for investigation are feasible, and

evaluating the competence of the applicant (Cole 1992). Fellowship or grant peer reviews – when compared to journal peer reviews – are perceived as entailing a heightened risk for judgments and decisions with low predictive validity. Accordingly, it is expected that studies on grant or fellowship peer reviews are less likely than studies on journal peer reviews to be able to confirm the predictive validity.

In recent years, besides the qualitative form of research evaluation, the peer review system, the quantitative form has become more and more important. “Measurement of research excellence and quality is an issue that has increasingly interested governments, universities, and funding bodies as measures of accountability and quality are sought” (Steele et al. 2006: 278). Weingart (2005a) notes that a really enthusiastic acceptance of bibliometric figures for evaluative purposes or for comparing the research success of scientists can be observed today. University rankings are normally based on bibliometric measures. The United Kingdom is planning to allocate government funding for research by universities in large part using bibliometric indicators: “The Government has a firm presumption that after the 2008 RAE [Research Assessment Exercise], the system for assessing research quality and allocating ‘quality-related’ (QR) research funding to universities from the Department for Education and Skills will be mainly metrics-based” (UK Office of Science and Technology 2006: 3). With the easy availability of bibliometric data and ready-to-use tools for generating bibliometric indicators for evaluation purposes, there is a danger of improper use.

As noted above, two competing theories of citing were developed in past decades: the normative theory of citing and the social constructive approach to citing. Following normative theory, the reasons why scientists cite documents are that the documents are relevant to their topic and provide useful background for their research and in order to acknowledge intellectual debt. The social constructive view on citing contradicts these assumptions. According to this view, citations are a social psychological process, not free of personal bias or social pressures and probably not made for the same reasons. While Cronin (1984) finds the existence of two competing theories of citing behavior hardly surprising, as the construction of scientific theory is generally characterized by ambivalence, for Liu (1997) and Weingart (2005b), the long-term oversimplification of thinking in terms of two theories reflects the absence of one satisfactory and accepted theory on which the better informed use of citation indicators could be based. Whereas Liu (1997) and Nicolaisen (2003) see the dynamic linkage of both theories as a necessary step in the quest for a satisfactory theory of citation, Garfield (1998) states: “There is no way to predict whether a particular citation (use of a reference by a new author) will be ‘relevant’” (p. 70).

The results of the studies presented in Sect. 8.3 suggest that not only the content of scientific work, but also other, in part non-scientific, factors play a role in citing. Citations can therefore be viewed as a complex, multi-dimensional and not a uni-dimensional phenomenon. The reasons authors cite can vary from scientist to scientist. On the basis of the available findings, should we then conclude that citation counts are not appropriate indicators of the impact of research? Are citation counts

not suitable for use in university rankings? Not so, says van Raan (2005a): “So undoubtedly the process of citation is a complex one, and it certainly not provides an ‘ideal’ monitor on scientific performance. This is particularly the case at a statistically low aggregation level, e.g., the individual researcher. There is, however, sufficient evidence that these reference motives are not so different or ‘randomly given’ to such an extent that the phenomenon of citation would lose its role as a reliable measure of impact. Therefore, application of citation analysis to the entire work, the ‘oeuvre’ of *a group of researchers as a whole over a longer period of time*, does yield in many situations a strong indicator of scientific performance” (p. 134–135, see also Laloë and Mosseri 2009).

Research on the predictive validity of peer review indicates that peer review is generally a credible method for evaluation of manuscripts and – in part – of grant and fellowship applications. But this overview of the reliability and fairness of the peer review process shows that there are also problems with peer reviews. However, despite its flaws, having scientists judge each other’s work is widely considered to be the “least bad way” to weed out weak work (Enserink 2001). In a similar manner, bibliometric indicators do have specific drawbacks. However, on a higher aggregation level (a larger group of scientists), it seems to be a reliable indicator of research impact. It has been frequently recommended that peer review should be used for the evaluation of scientific work and should be supplemented with bibliometrics (and other metrics of science) to yield a broader and powerful methodology for assessment of scientific advancement (Geisler 2001; van Raan 1996). Thus, the combination of both indicators in university rankings seems to be a sensible way to build on the strengths and compensate for the weaknesses of both evaluative instruments.

References

- Abelson, P. H. (1980). Scientific communication. *Science*, 209(4452), 60–62.
- Armstrong, P. W., Caverson, M. M., Adams, L., Taylor, M., & Olley, P. M. (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.
- Ayres, I., & Vars, F. E. (2000). Determinants of citations to articles in elite law reviews. *Journal of Legal Studies*, 29(1), 427–450.
- Bar-Ilan, J. (2009). A closer look at the sources of informetric research. *Cybermetrics*, 13(1), p. 4.
- Bartley, W. W. (1984). *The retreat to commitment* (2nd ed.). La Salle: Open Court.
- Bayer, A. E., & Folger, J. (1966). Some correlates of a citation measure of productivity in science. *Sociology of Education*, 39(4), 381–390.
- Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics [Review]. *Annual Review of Information Science and Technology*, 36, 3–72.
- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45, 199–245.
- Bornmann, L., & Daniel, H.-D. (2005a). Criteria used by a peer review committee for selection of research fellows – a boolean probit analysis. *International Journal of Selection and Assessment*, 13(4), 296–303.

- Bornmann, L., & Daniel, H.-D. (2005b). Selection of research fellowship recipients by committee peer review. Analysis of reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics*, *63*(2), 297–320.
- Bornmann, L., & Daniel, H.-D. (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.
- Bornmann, L., & Daniel, H.-D. (2008a). The effectiveness of the peer review process: Inter-referee agreement and predictive validity of manuscript refereeing at *Angewandte Chemie. Angewandte Chemie. International Edition*, *47*(38), 7173–7178. doi:10.1002/anie.200800513.
- Bornmann, L., & Daniel, H.-D. (2008b). 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. doi:10.1002/asi.20901.
- Bornmann, L., & Daniel, H.-D. (2008c). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. doi:10.1108/00220410810844150.
- Bornmann, L., & Daniel, H.-D. (2009a). The luck of the referee draw: The effect of exchanging reviews. *Learned Publishing*, *22*(2), 117–125. doi:10.1087/2009207.
- Bornmann, L., & Daniel, H.-D. (2009b). Universality of citation distributions. A validation of Radicchi et al.'s relative indicator $c_i = c/c_0$ at the micro level using data from chemistry. *Journal of the American Society for Information Science and Technology*, *60*(8), 1664–1670.
- Bornmann, L., & Daniel, H.-D. (2010). The manuscript reviewing process – empirical research on review requests, review sequences and decision rules in peer review. *Library & Information Science Research*, *32*(1), 5–12. doi:10.1016/j.lisr.2009.07.010.
- Bornmann, L., Wallon, G., & Ledin, A. (2008). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European Molecular Biology Organization programmes. *PLoS One*, *3*(10), e3480.
- Bornmann, L., Marx, W., Schier, H., Thor, A., & Daniel, H.-D. (2010). From black box to white box at open access journals: Predictive validity of manuscript reviewing and editorial decisions at *Atmospheric Chemistry and Physics. Research Evaluation*, *19*(2), 81–156.
- Bornstein, R. F. (1991). The predictive validity of peer-review: A neglected issue. *The Behavioral and Brain Sciences*, *14*(1), 138–139.
- Boyack, K. W., & Klavans, R. (2005). Predicting the importance of current papers. In P. Ingwersen & B. Larsen (Eds.), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (Vol. 1, pp. 335–342). Stockholm: Karolinska University Press.
- Buela-Casal, G., Gutiérrez-Martínez, O., Bermúdez-Sánchez, M., & Vadillo-Muñoz, O. (2007). Comparative study of international academic rankings of universities. *Scientometrics*, *71*(3), 349–365.
- Butler, D. (2007). Academics strike back at spurious rankings. *Nature*, *447*(7144), 514–515.
- Carpenter, M. P., & Narin, F. (1981). The adequacy of the Science Citation Index (SCI) as an indicator of international scientific activity. *Journal of the American Society for Information Science*, *32*(6), 430–439.
- Carter, G. (1982). *What we know and do not know about the peer review system* (Rand Report N-1878-RC/NIH). Santa Monica: RAND Corporation.
- Castellano, C., & Radicchi, F. (2009). On the fairness of using relative indicators for comparing citation performance in different disciplines [Article]. *Archivum Immunologiae Et Therapiae Experimentalis*, *57*(2), 85–90. doi:10.1007/s00005-009-0014-0.
- Cole, S. (1992). *Making science: Between nature and society*. Cambridge: Harvard University Press.
- Cole, J. R., & Cole, S. (1973). *Social stratification in science*. Chicago: The University of Chicago Press.
- Collins, H. (2004). *Gravity's shadow: The search for gravitational waves*. Chicago: The University of Chicago Press.

- Cozzens, S. E. (1985). Comparing the sciences – Citation context analysis of papers from neuropharmacology and the sociology of science. *Social Studies of Science*, 15(1), 127–153.
- Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. Oxford: Taylor Graham.
- Cronin, B. (2005). *The hand of science*. Academic Writing and its Rewards, Scarecrow Press, Lanham, MD.
- Daniel, H.-D. (1993). *Guardians of science: Fairness and reliability of peer review*. Weinheim: Wiley-VCH.
- Daniel, H.-D. (2005). Publications as a measure of scientific advancement and of scientists' productivity. *Learned Publishing*, 18, 143–148.
- de Vries, D. R., Marschall, E. A., & Stein, R. A. (2009). Exploring the peer review process: What is it, does it work, and can it be improved? *Fisheries*, 34(6), 270–279.
- Eckes, T. (2004). Rater agreement and rater severity: A many-faceted Rasch analysis of performance assessments in the "Test Deutsch als Fremdsprache" (TestDaF). *Diagnostica*, 50(2), 65–77.
- Ehse, I. (2004). By scientists, for scientists. The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – and how it functions. *B.I.F. Futura*, 19, 170–177.
- Eichorn, P., & Yankauer, A. (1987). Do authors check their references – a survey of accuracy of references in 3 public-health journals. *American Journal of Public Health*, 77(8), 1011–1012.
- Eisenhart, M. (2002). The paradox of peer review: Admitting too much or allowing too little? *Research in Science Education*, 32(2), 241–255.
- Enserink, M. (2001). Peer review and quality: A dubious connection? *Science*, 293(5538), 2187–2188.
- Enserink, M. (2007). Who ranks the university rankers? *Science*, 317(5841), 1026–1028. doi:10.1126/science.317.5841.1026.
- Ernst, E., Saradeth, T., & Resch, K. L. (1993). Drawbacks of peer review. *Nature*, 363(6427), 296.
- Evans, J. T., Nadjari, H. I., & Burchell, S. A. (1990). Quotational and reference accuracy in surgical journals – A continuing peer-review problem. *Journal of the American Medical Association*, 263(10), 1353–1354.
- Feist, G. J. (2006). *The psychology of science and the origins of the scientific mind*. New Haven: Yale University Press.
- Figueredo, E. (2006). The numerical equivalence between the impact factor of journals and the quality of the articles. *Journal of the American Society for Information Science and Technology*, 57(11), 1561.
- Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper – Diversity and uniqueness in reviewer comments. *The American Psychologist*, 45(5), 591–598.
- Fok, D., & Franses, P. H. (2007). Modeling the diffusion of scientific publications. *Journal of Econometrics*, 139(2), 376–390. doi:10.1016/j.jeconom.2006.10.021.
- Fox, T. (1965). *Crisis in communication: The functions and future of medical publication*. London: Athlone Press.
- Garfield, E. (1970). Citation indexing for studying science. *Nature*, 227(5259), 669–671.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060), 471–479.
- Garfield, E. (1998). Random thoughts on citationology. Its theory and practice [Article]. *Scientometrics*, 43(1), 69–76.
- Garfield, E. (2002). Highly cited authors. *Scientist*, 16(7), 10.
- Garfield, E., & Welljamsdorff, A. (1992). Citation data – Their use as quantitative indicators for science and technology evaluation and policy-making. *Current Contents*, 49, 5–13.
- Geisler, E. (2001). The mires of research evaluation. *The Scientist*, 15(10), 39.
- Godlee, F., & Dickersin, K. (2003). Bias, subjectivity, chance, and conflict of interest. In F. Godlee & J. Jefferson (Eds.), *Peer review in health sciences* (2nd ed., pp. 91–117). London: BMJ Publishing Group.
- Goodman, S. N., Berlin, J., Fletcher, S. W., & Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Annals of Internal Medicine*, 121(1), 11–21.

- Hames, I. (2007). *Peer review and manuscript management of scientific journals: Guidelines for good practice*. Oxford: Blackwell.
- Hemlin, S. (1996). Research on research evaluations. *Social Epistemology*, 10(2), 209–250.
- Hojat, M., Gonnella, J. S., & Caelleigh, A. S. (2003). Impartial judgment by the “gatekeepers” of science: Fallibility and accountability in the peer review process. *Advances in Health Sciences Education*, 8(1), 75–96.
- Hornbostel, S., Böhmer, S., Klingsporn, B., Neufeld, J., & von Ins, M. (2009). Funding of young scientist and scientific excellence. *Scientometrics*, 79(1), 171–190.
- Hudson, J. (2007). Be known by the company you keep: Citations – quality or chance? *Scientometrics*, 71(2), 231–238.
- Hurt, C. D. (1987). Conceptual citation differences in science, technology, and social sciences literature. *Information Processing and Management*, 23(1), 1–6.
- Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2001). Peer review in the funding of research in higher education: The Australian experience. *Educational Evaluation and Policy Analysis*, 23(4), 343–346.
- Jensen, P., Rouquier, J. B., & Croissant, Y. (2009). Testing bibliometric indicators by their prediction of scientists promotions. [Article]. *Scientometrics*, 78(3), 467–479. doi:10.1007/s11192-007-2014-3.
- Judge, T., Cable, D., Colbert, A., & Rynes, S. (2007). What causes a management article to be cited – article, author, or journal? *The Academy of Management Journal (AMJ)*, 50(3), 491–506.
- Kellogg, D. (2006). Toward a post-academic science policy: Scientific communication and the collapse of the Mertonian norms. *International Journal of Communications Law & Policy*, 11, IJCLP Web-Doc 1-11-2006.
- Kellsey, C., & Knievel, J. E. (2004). Global English in the humanities? A longitudinal citation study of foreign-language use by humanities scholars. *College & Research Libraries*, 65(3), 194–204.
- King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science*, 13(5), 261–276.
- Knorr-Cetina, K. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon.
- Knorr-Cetina, K. (1991). Merton sociology of science: The first and the last sociology of science. *Contemporary Sociology*, 20(4), 522–526.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- LaFollette, M. C. (1992). *Stealing into print: Fraud, plagiarism and misconduct in scientific publishing*. Berkeley: University of California Press.
- Laloë, F., & Mosseri, R. (2009). Bibliometric evaluation of individual researchers: Not even right... not even wrong! *Europhysics News*, 40(5), 26–29.
- Langfeldt, L. (2006). The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation*, 15(1), 31–41.
- Lansingh, V. C., & Carter, M. J. (2009). Does Open Access in ophthalmology affect how articles are subsequently cited in research? [Article]. *Ophthalmology*, 116(8), 1425–1431. doi:10.1016/j.ophtha.2008.12.052.
- Larivière, V., & Gingras, Y. (2010). The impact factor’s Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology*, 61(2), 424–427.
- Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. London: Sage.
- Lawani, S. M. (1977). The professional literature used by American and French agronomists and the implications for agronomic education. *Journal of Agronomic Education*, 6, 41–46.
- Lawani, S. M. (1986). Some bibliometric correlates of quality in scientific research. *Scientometrics*, 9(1–2), 13–25.
- Lawrence, S. (2001). Free online availability substantially increases a paper’s impact. *Nature*, 411(6837), 521–521.

- Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution*, 20(1), 28–32.
- Lewison, G. (1998). Gastroenterology research in the United Kingdom: Funding sources and impact. *Gut*, 43(2), 288–293.
- Lienert, G. A. (1987). *Schulnoten-evaluation*. Frankfurt am Main: Athenäum.
- Lindsey, D. (1989). Using citation counts as a measure of quality in science. Measuring what's measurable rather than what's valid. *Scientometrics*, 15(3–4), 189–203.
- Liu, Z. M. (1997). Citation theories in the framework of international flow of information: New evidence with translation analysis. *Journal of the American Society for Information Science*, 48(1), 80–87.
- Long, J. S., & Fox, M. F. (1995). Scientific careers – Universalism and particularism. *Annual Review of Sociology*, 21, 45–71.
- Lundberg, J. (2007). Lifting the crown – citation z-score. *Journal of Informetrics*, 1(2), 145–154.
- Mählck, P., & Persson, O. (2000). Socio-bibliometric mapping of intra-departmental networks. *Scientometrics*, 49(1), 81–91.
- Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science Technology Human Values*, 34(5), 573–606. doi:10.1177/0162243908329381.
- Martin, B. R., & Irvine, J. (1983). Assessing basic research – Some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12(2), 61–90.
- McClellan, J. E. (2003). *Specialist control – The publications committee of the Academie Royal des Sciences (Paris) 1700–1793* (Transactions of the American Philosophical Society, Vol. 93). Philadelphia: American Philosophical Society.
- McDonald, J. D. (2007). Understanding journal usage: A statistical analysis of citation and use. *Journal of the American Society for Information Science and Technology*, 58(1), 39–50.
- McDonald, R. J., Cloft, H. J., & Kallmes, D. F. (2009). Fate of manuscripts previously rejected by the *American Journal of Neuroradiology*: A follow-up analysis. *American Journal of Neuroradiology*, 30(2), 253–256. doi:10.3174/Ajnr.A1366.
- Melin, G., & Danell, R. (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.
- Menou, M. J. (1983). Cultural barriers to the international transfer of information. *Information Processing and Management*, 19(3), 121–129.
- Merisotis, J., & Sadlak, J. (2005). Higher education rankings: Evolution, acceptance, and dialogue. *Higher Education in Europe*, 30(2), 97–101.
- Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1, 115–126.
- Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review*, 22(6), 635–659. doi:10.2307/2089193.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
- Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *ISIS*, 79(4), 606–623.
- Moed, H. (2008). UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics*, 74(1), 153–161.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill: Computer Horizons.
- National Academy of Sciences. (2006). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. Washington: The National Academies Press.
- Nicolaisen, J. (2002). The J-shaped distribution of citedness. *Journal of Documentation*, 58(4), 383–395.
- Nicolaisen, J. (2003). The social act of citing: Towards new horizons in citation theory. In *The social act of citing: towards new horizons in citation theory. Proceedings of the 66th ASIST Annual Meeting*, 12–20.

- Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, 41, 609–641.
- Oppenheim, C. (1995). The correlation between citation counts and the 1992 research assessment exercise ratings for British library and information science university departments. *Journal of Documentation*, 51(1), 18–27.
- Oppenheim, C. (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.
- Ophthof, T., Furstner, F., van Geer, M., & Coronel, R. (2000). Regrets or no regrets? No regrets! The fate of rejected manuscripts. *Cardiovascular Research*, 45(1), 255–258.
- Owen, R. (1982). Reader bias. *Journal of the American Medical Association*, 247(18), 2533–2534.
- Peters, H. P. F., & van Raan, A. F. J. (1994). On determinants of citation scores – A case study in chemical engineering. *Journal of the American Society for Information Science*, 45(1), 39–49.
- Pierie, J. P. E. N., Walvoort, H. C., & Overbeke, A. J. P. M. (1996). Readers' evaluation of effect of peer review and editing on quality of articles in the *Nederlands Tijdschrift voor Geneeskunde*. *Lancet*, 348(9040), 1480–1483.
- Popper, K. R. (1961). *The logic of scientific discovery* (2nd ed.). New York: Basic Books.
- Pruthi, S., Jain, A., Wahid, A., Mehra, K., & Nabi, S. A. (1997). Scientific community and peer review system – A case study of a central government funding scheme in India. *Journal of Scientific and Industrial Research*, 56(7), 398–407.
- Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(45), 17268–17272. doi:10.1073/pnas.0806977105.
- Reinhart, M. (2009). Peer review of grant applications in biology and medicine. Reliability, fairness, and validity. *Scientometrics*, 81(3), 789–809. doi:10.1007/s11192-008-2220-7.
- Research Evaluation and Policy Project. (2005). *Quantitative indicators for research assessment – A literature review* (REPP discussion paper 05/1). Canberra: Research Evaluation and Policy Project, Research School of Social Sciences, The Australian National University.
- Ross, P. F. (1980). *The sciences' self-management: Manuscript refereeing, peer review, and goals in science*. Lincoln: The Ross Company.
- Sandström, U., Wadskog, D., & Karlsson, S. (2005). Research institutes and universities: Does collaboration pay? In P. Ingwersen & B. Larsen (Eds.), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (Vol. 2, pp. 690–691). Stockholm, Sweden: Karolinska University Press.
- Schmelkin, L. (2006). *Weaknesses of peer reviewing and peer refereeing*. Paper presented at the first international conference on Knowledge Communication and Peer Reviewing, Orlando.
- Schneider, J. W. (2009). An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science*, 8(3), 364–378. doi:10.1057/Eps.2009.19.
- Shadbolt, N., Brody, T., Carr, L., & Harnad, S. (2006). The Open Research Web: A preview of the optimal and the inevitable. In N. Jacobs (Ed.), *Open access: Key strategic, technical and economic aspects* (pp. 195–208). Oxford: Chandos.
- Sharp, D. W. (1990). What can and should be done to reduce publication bias? The perspective of an editor. *Journal of the American Medical Association*, 263(10), 1390–1391.
- Shashok, K. (2005). Standardization vs diversity: How can we push peer review research forward? *Medscape General Medicine*, 7(1), 11.
- Shatz, D. (2004). *Peer review: A critical inquiry*. Lanham: Rowman & Littlefield.
- Sismondo, S. (1993). Some social constructions [Article]. *Social Studies of Science*, 23(3), 515–553.
- Smith, L. C. (1981). Citation analysis. *Library Trends*, 30(1), 83–106.
- Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. [Article]. *Journal of the Royal Society of Medicine*, 99(4), 178–182.

- Stamps, A. E. (1997). Advances in peer review research: An introduction. *Science and Engineering Ethics*, 3(1), 3–10.
- Steele, C., Butler, L., & Kingsley, D. (2006). The Publishing imperative: The pervasive influence of publication metrics. *Learned Publishing*, 19(4), 277–290.
- Stewart, J. A. (1983). Achievement and ascriptive processes in the recognition of scientific articles. *Social Forces*, 62(1), 166–189.
- Tijssen, R. J. W., van Leeuwen, T. N., & van Raan, A. F. J. (2002). *Mapping the scientific performance of German medical research. An international comparative bibliometric study*. Stuttgart: Schattauer.
- UK Office of Science and Technology. (2006). *Science and innovation investment framework 2004–2014: Next steps*. London: UK Office of Science and Technology.
- Unver, B., Senduran, M., Kocak, F. U., Gunal, I., & Karatosun, V. (2009). Reference accuracy in four rehabilitation journals [Article]. *Clinical Rehabilitation*, 23(8), 741–745. doi:10.1177/0269215508102968.
- van den Besselaar, P., & Leydesdorff, L. (2007). *Past performance as predictor of successful grant applications: A case study*. Den Haag: Rathenau Instituut.
- van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, 36(3), 397–420.
- van Raan, A. F. J. (2005a). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133–143.
- van Raan, A. F. J. (2005b). For your citations only? Hot topics in bibliometric analysis. *Measurement: Interdisciplinary Research and Perspectives*, 3(1), 50–62.
- Weingart, P. (2005a). Das Ritual der Evaluierung und die Verführbarkeit. In P. Weingart (Ed.), *Die Wissenschaft der Öffentlichkeit: Essays zum Verhältnis von Wissenschaft, Medien und Öffentlichkeit* (pp. 102–122). Weilerswist: Velbrück.
- Weingart, P. (2005b). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62(1), 117–131.
- Weller, A. C. (2002). *Editorial peer review: Its strengths and weaknesses*. Medford: Information Today.
- White, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52(2), 87–108.
- Whitley, R., & Gläser, J. (Eds.). (2007). *The changing governance of the sciences: The advent of research evaluation systems*. Dordrecht: Springer.
- Wiley, S. (2008). Peer review isn't perfect ... But it's not a conspiracy designed to maintain the status quo. *The Scientist*, 22(11), 31.
- Wilson, J. D. (1978). Peer review and publication. *The Journal of Clinical Investigation*, 61(4), 1697–1701.
- Wood, F. Q., & Wessely, S. (2003). Peer review of grant applications: A systematic review. In F. Godlee & T. Jefferson (Eds.), *Peer review in health sciences* (2nd ed., pp. 14–44). London: BMJ Books.
- Ziman, J. (2000). *Real science: What it is, and what it means*. Cambridge: Cambridge University Press.