

Qualität und Wirkung - Von der guten zur signifikanten Forschung

Workshop der Österreichischen Forschungsgemeinschaft
5. - 6. 6. 2009

Lässt sich Qualität von Forschung messen?

Dr. Lutz Bornmann

ETH Zurich

Professorship for Social Psychology and Research on Higher Education

Zähringerstr. 24, CH-8092 Zurich

bornmann@gess.ethz.ch

Grundsätzlich können wir bei Bewertungen in der Wissenschaft zwischen einer ‚qualitativen‘ Form, der Bewertung einer wissenschaftlichen Arbeit (z. B. eines Manuskripts oder Forschungsantrags) durch kompetente Peers (Hornbostel, 1997), und einer ‚quantitativen‘ Form, der Bewertung von wissenschaftlicher Arbeit anhand bibliometrischer Indikatoren, unterscheiden (Daniel, 2001).

Bewertungen in der Wissenschaft durch Peer Review

Mit der Gründung wissenschaftlicher Gesellschaften – insbesondere der Royal Society (London) in England – entstanden Mitte des 17. Jahrhundert erste Formen des Peer Review, mit denen die Güte der Manuskripte, die in den Zeitschriften der Gesellschaften veröffentlicht wurden, sichergestellt werden sollte (Biagioli, 2002; McClellan, 2003). Die Gesellschaften waren mit dem Problem konfrontiert, dass Personen, die keine Mitglieder und damit für eine Gesellschaft nicht vertrauenswürdig waren, mit Ergebnissen von Experimenten vorstellig wurden, über deren Güte wenig Klarheit bestand. Um die Vertrauenswürdigkeit in die Ergebnisse herzustellen, ließen die Gesellschaften es entweder zu, dass ein Mitglied für die Richtigkeit der Ergebnisse bürgte (nach Shapiro, 2002, wurde dieses Vorgehen aus der Jurisdiktion übernommen), oder mehrere Mitglieder der Gesellschaft nahmen eine kritische Begutachtung der wissenschaftlichen Arbeit vor. Dabei ließen sie anfangs Experimente vor ihren Augen wiederholen, später begutachteten sie die Beschreibung der Durchführung und der Resultate von Experimenten in Manuskripten.

Peer Review Verfahren sind in der Regel so gestaltet, dass institutionell beauftragte Wissenschaftler ein ‚Evaluationsobjekt‘ prüfen und legitimieren, und der gutachterliche Bericht dann Entscheidungsträgern zur weiteren Verwendung vorgelegt wird (z. B. für die Annahme oder die Ablehnung eines Manuskriptes oder Forschungsantrags). Je nach Anwendungsbereich (aber auch innerhalb eines Anwendungsbereiches) kann die *konkrete* Ausgestaltung des Begutachtungsprozesses mehr oder weniger stark variieren (siehe dazu Hansson, 2002; Marsh & Ball, 1991; Shashok, 2005). Die Begutachtung kann „nach hoch formalisierten Vorgaben erfolgen, die Auswahl bestimmter Urteilkriterien kann aber auch den Gutachtern (Peers) überlassen bleiben“ (Hornbostel, 1997, S. 195). Gutachter können anonym oder autonym urteilen. Die Begutachteten können anonymisiert sein oder für die Gutachter ‚sichtbar‘ sein (‚double-blind‘ versus ‚single-blind‘). Ein Gutachter kann permanent oder ad hoc eingesetzt werden (Geisler, 2000). Eine Begutachtung kann von einem Gutachter oder von mehreren Gutachtern durchgeführt werden (Marsh & Ball, 1991). Wird eine Gruppe von Gutachtern eingesetzt, können die Mitglieder unabhängig oder abhängig voneinander begutachten (Geisler, 2000). Der Peer Review Prozess kann öffentlich einsehbar oder vertraulich konzipiert sein.

Die typischen Anwendungsbereiche des Peer Review in der Wissenschaft ist die Begutachtung von Manuskripten und Forschungsanträgen. In diesen Bereichen haben die Gutachter die Aufgabe, die ‚beste‘ wissenschaftliche Forschung unter der Bedingung knapper Ressourcen (d. h. einer begrenzten Anzahl an Druckseiten in Zeitschriften bzw. begrenzten Forschungsmitteln) zur Auswahl zu empfehlen und gegebenenfalls Verbesse-

rungsvorschläge zu den eingereichten Arbeiten zu formulieren (Enserink, 2001; Hackett & Chubin, 2003). Darüber hinaus sollen die Gutachter Fehler in wissenschaftlichen Arbeiten oder wissenschaftliches Fehlverhalten im Zusammenhang mit den Arbeiten aufdecken (Smith, 2006). „Peer review should be a filter to separate the wheat from the chaff, and thus to guarantee and improve the quality of published research“ (Grivell, 2006, S. 567).

Von den Befürwortern des Peer Review werden aktive Wissenschaftler als die am besten geeigneten Personen angesehen, die Arbeit ihrer Fachkollegen auf ihre wissenschaftliche Güte hin zu begutachten (Eisenhart, 2002; Geisler, 2000). Peer Review würde wie kein anderes Verfahren „eine wissenschaftsadäquate Bewertung der Leistungen sowie die verfassungsrechtlich notwendige Einbeziehung des wissenschaftlichen Sachverständes“ (Daniel, 2006, S. 185) gewährleisten. „The peer review system has for a long time been highly effective to increase the quality of published work. We have enormously profited from the altruism of unpaid first-class scientists who devoted their time to the improvement of the science of others“ (Campanario & Acedo, 2007, S. 740). Nach Graefen und Thielmann (2007) trägt die Situation, dass Wissenschaftler in Peer Review Verfahren sowohl als Gutachtende als auch als Begutachtete auftreten „zur Vereinheitlichung von Bewertungskriterien wie auch zur Verallgemeinerung formaler Anforderungen“ (S. 69) bei.

Die Kritiker des Peer Review (siehe z. B. Abate, 1995; Finn, 2002; Horrobin, 2001; Ross, 1980; Roy, 1985) sehen vor allem drei Schwächen des Verfahrens: (1) Verschiedene Gutachter würden kaum in der Bewertung ein und derselben wissenschaftlichen Arbeit übereinstimmen (fehlende Reliabilität). (2) Gutachterliche Empfehlungen würden systematische Urteilsfehler aufweisen. Die Urteile würden nicht nur auf der wissenschaftlichen Qualität einer Arbeit, sondern auch auf nicht-wissenschaftlichen Kriterien basieren (fehlende Fairness). (3) Dementsprechend sei der Zusammenhang zwischen den Urteilen im Peer Review und der Qualität der begutachteten Arbeit gering (fehlende Validität). Der einzige Grund – so die Kritiker – für die weitere Verwendung des Peer Review Verfahrens sei, dass es keinen klaren Konsens über eine ‚bessere‘ Alternative gäbe (Young, 2003).

Bewertungen in der Wissenschaft durch bibliometrische Indikatoren

Neben die ex ante und ex post Bewertung von wissenschaftlicher Arbeit durch Fachgutachter ist seit Ende der 1980er Jahre zunehmend die ex post Bewertung anhand bibliometrischer Indikatoren getreten. Bei dieser Form der Bewertung werden Publikations- und Zitierhäufigkeiten als numerisches Relativ für die Güte einer wissenschaftlichen Arbeit verwendet (Hornbostel, 1997). In einer Zeit, in der Evaluationen einen immer größer werdenden Einfluss auf die Steuerung der Wissenschaft ausüben, kann der Einsatz bibliometrischer Indikatoren zu einer Entlastung des Gutachterwesens führen (siehe dazu Neidhardt, 2006). Weingart (2005a) sieht dagegen als Grund für die zunehmende Popularität bibliometrischer Indikatoren die wachsende Skepsis gegenüber dem traditionellen Peer Review Verfahren. Von den Indikatoren wird eine Objektivierung von Bewertungsprozessen erwartet (Adam, 2002; Hanlin, 1984). Publikationen und Zitierungen werden als ‚attraktive‘ Rohdaten angesehen, da „[they] do not require the cooperation of a respondent and do not themselves contaminate the response (i.e., they are non-reactive)“ (Smith, 1981, S. 84).

Die Anregung zur Konstruktion bibliometrischer Indikatoren für Bewertungen in der Wissenschaft ging von dem amerikanischen Wissenschaftshistoriker de Solla Price (1965) aus. Die Verwendung von Publikationshäufigkeiten als bibliometrischer Indikator beruht darauf, dass Wissenschaftler ihre Zwischen- und Endresultate für gewöhnlich in wissenschaftlichen Arbeiten publizieren (Mulligan & Mabe, 2006; Smith, 1988), denn „research results do not count as scientific unless they are reported, disseminated, shared, and eventually transformed into communal property, by being formally *published* [Hervorhebung durch den Autor]“ (Ziman, 2000, S. 110). Die Anzahl an Publikationen, die ein Wissenschaftler über einen bestimmten Zeitraum veröffentlicht hat, kann demnach als ein numerisches Relativ dafür angesehen werden, wie produktiv er in dieser Zeit in der Forschung gewesen ist.

Die Publikation einer wissenschaftlichen Arbeit dient jedoch nicht nur der Verbreitung von Zwischen- und Endresultaten. Sie ist auch als ‚Aufforderung‘ für andere Wissenschaftler gedacht, die beschriebenen Befunde für die eigene Forschung zu verwenden (McClellan, 2003). Kommen Wissenschaftler der Aufforderung zur Verwendung nach, kennzeichnen sie dies in ihren eigenen Publikationen durch einen Verweis. Da ein Verweis anzeigen soll, dass in eine Publikation Inhalte einer anderen Publikation (d. h. Forschungsergebnisse, Gedankengut etc.) Eingang gefunden haben (McClellan, 2003), werden Zitierhäufigkeiten (d. h. die Anzahl der Verweise) in der Wissenschaftsevaluation als ein Indikator für die *Wirkung* von Forschung verwendet: „The *impact* [Hervorhebung durch die Autoren] of a piece of research is the degree to which it has been useful to other researchers and users“ (Shadbolt, Brody, Carr & Harnad, 2006, S. 202). Als Indikator für die Wirkung publizierter wissenschaftlicher Arbeiten würden Zitierungen nach van Raan (2004) einen wichtigen Teilaspekt von wissenschaftlicher Qualität messen: „Citation-based bibliometric analysis provides indicators of international impact, influence. This can be regarded as, at least, one crucial aspect of scientific quality, and thus a ‚proxy‘ of quality as follows from a long standing experience in bibliometric analysis“ (S. 27, siehe auch Martin & Irvine, 1983). Nach dem Research Evaluation and Policy Project (2005) „there is an emerging trend in science policy to regard *impact*, the measurable part of quality, as a *proxy measure for quality* in total [Hervorhebungen durch den Autor]“ (S. 4).

In bibliometrischen Datenbanken, wie dem Science Citation Index (SCI) von Thomson Scientific (Philadelphia, PA, USA), werden neben den bibliographischen Angaben einer Veröffentlichung deren Zitierungen in anderen Publikationen erfasst (für einen Überblick über die wichtigsten bibliometrischen Datenbanken siehe Meho, 2007; Neuhaus & Daniel, 2008). In der Wissenschaftsevaluation werden diese Daten für die Bewertung von wissenschaftlicher Arbeit mit (mehr oder weniger) standardisierten Verfahren ausgewertet (Weingart, 2001). Waren in der Vergangenheit ausschließlich Experten in der Lage, mit diesen Daten zu arbeiten, bietet Thomson Scientific seit kurzem „gebrauchsfertige ISI-Indikatorenpakete“ (Weingart, 2005a, S. 107) an, mit denen auch Laien ohne große Mühe bibliometrische Kennzahlen erzeugen können (siehe auch Steele, Butler & Kingsley, 2006).

Während die Verwendung von Publikationshäufigkeiten als Indikator für die Produktivität von Wissenschaftlern heutzutage weitgehend anerkannt ist, wird gegen die Verwendung von Zitierhäufigkeiten als Indikator für die Qualität einer wissenschaftlichen Arbeit seit Jahren eine Vielzahl von Bedenken geäußert. Zitierhäufigkeiten seien keine ‚objektiven‘ Messungen von wissenschaftlicher Qualität, sondern ein kritisierbares Messkonstrukt (Hornbostel, 1997). So wird unter anderem kritisiert, dass (1) wissenschaftliche Qualität ein komplexes Phänomen darstelle, das nicht auf einer eindimensionalen Skala (d. h. anhand von Zitierhäufigkeiten) gemessen werden könne (siehe dazu z. B. die Arbeiten von Barbui, Cipriani, Malvini & Tansella, 2006; Berghmans et al., 2003). „The quality of publications cannot be measured using citation indices, whatever their statistical reliability“ (Döring, 2007). (2) Von den Autoren würden Publikationen zitiert werden, die den Inhalt ihrer Arbeiten kaum intellektuell bzw. kognitiv beeinflusst haben. (3) Viele Publikationen würden von Autoren nicht zitiert werden, obwohl sie deren Arbeiten intellektuell bzw. kognitiv beeinflusst haben. (4) Autoren würden relevante Publikationen übersehen (Wright & Armstrong, 2007).

Die Verknüpfung von bibliometrischen Indikatoren und Peer Review

Als ein Grund für die Popularität, die bibliometrische Indikatoren bei Politikern aus den Wissenschaftsressorts genießen, wird deren Skepsis gegenüber dem Peer Review, d. h. der Sicherung von Qualität durch Kollegen aus dem gleichen Fachgebiet, gesehen. „Das Vertrauen, das den *peer review*-Mechanismen [Hervorhebung durch den Autor] entzogen wird, wird auf die numerischen Indikatoren verlagert“ (Weingart, 2005a, S. 110). Von bibliometrischen Indikatoren könne man nicht nur unabhängiger und objektiver Bewertungen als von Fachgutachtern erwarten, sie seien auch billiger und weit weniger zeitaufwendig in der Anwendung (Kornhuber, 1988). Ferner würden in einem Wissenschaftssystem, das immer komplexer gestaltet sei, bibliometrische Analysen die

Auswertung von ‚Massendaten‘ erlauben, die für einzelne Gutachter nicht mehr zugänglich seien (siehe dazu Butler, 2004; Hornbostel, 1997).

Allerdings wird bei dieser Gegenüberstellung von Peer Review und bibliometrischen Indikatoren häufig nicht bedacht, dass beide Instrumente eng miteinander verknüpft sind (Hornbostel, 1997). „Editorial peer review, and its quality and fairness, ... have a central role in all other assessments“ (Grivell, 2006, S. 567). Die bibliometrischen Indikatoren resultieren aus einem wissenschaftlichen Kommunikationsprozess, der sich über die Veröffentlichungen in Fachzeitschriften entfaltet, womit sie auch die Begutachtungen im Peer Review repräsentieren, die zu den Publikationen geführt haben. „Aus diesem Grund können sie nicht exakter oder objektiver sein, als diese [Begutachtungen im Peer Review] selbst“ (Weingart, 2005a, S. 111).

Während die einen das ‚qualitative‘ Peer Review und die ‚quantitativen‘ Techniken der Bibliometrie in Konkurrenz zueinander sehen (Bozeman, Dietz & Gaughan, 2001), empfiehlt die große Mehrheit der anderen, „dass sie sich wechselseitig ergänzen sollten, um die Schwächen des jeweils anderen Verfahrens zu kompensieren“ (Daniel, 2001, S. 11; siehe auch Hornbostel, 1997; van Raan, 1999).

Auf der einen Seite könne – so wird argumentiert – das Peer Review durch bibliometrische Indikatoren sowohl (1) informiert, als auch (2) einer Überprüfung unterzogen werden. (1) „Bibliometric analysis will never be a substitute for peer review, but, if the analysis is comprehensive and sound, it should inform peer review“ (Jones & Sizer, 1990, S. 321). Die Verwendung der Indikatoren könne Idiosynkrasien im Peer Review entgegenwirken, „weil sie [die Indikatoren] – wenn im Widerspruch zum Urteil der Gutachter stehend – die Explikation von Qualitätskriterien und die Begründung von Urteilen anregen“ (Gläser, 2006, S. 352). „The experience at INSERM [The French National Institute of Health and Medical Research, Paris] shows that introducing bibliometric indicators ... strengthens the transparency of decisions and provides for their substantiation“ (European Science Foundation, 2007, S. 20). (2) Für Weingart (2005a) besteht eine wichtige Funktion der Indikatoren darin, die Begutachtungen im Peer Review einer Überprüfung zu unterziehen, um sie besser gegen ein grundlegendes Problem des Peer Review – die Wirkungsweise der Vetterwirtschaft („old boys networks“) – zu schützen (Weingart, 2005b).

Auf der anderen Seite könnten die Ergebnisse bibliometrischer Analysen durch Gutachter kommentiert und interpretiert werden. Schneider (1988) hält eine „unmittelbare Bewertung von Forschungsleistungen nach *formalisierten, kontextunabhängigen* und *ohne zusätzliche Interpretationsleistungen* [Hervorhebungen durch den Autor] anwendbaren Kriterien“ (S. 440) für unmöglich. Es würden Fachgutachter benötigt, die eine situationsadäquate Relevanzspezifizierung vornehmen, indem sie Rückwirkungen auf Individuen und Institutionen berücksichtigen (Weingart, 2005a). Erst durch die Interpretation von Experten ließen sich aus den Ergebnissen bibliometrischer Analysen qualitative Einschätzungen wissenschaftlicher Leistungen entwickeln. Dabei bedürften die Indikatoren nicht nur der fachkundigen Kommentierung, sondern auch der Anreicherung und Korrektur durch Kontextwissen (Hornbostel, 1997).

Alles in allem gibt es demnach gute Gründe, das ‚qualitative‘ Peer Review und die ‚quantitativen‘ Techniken der Bibliometrie nicht als konkurrierende Verfahren in der Wissenschaftsevaluation anzusehen, sondern die Anwendung beider Verfahren als Möglichkeit zu betrachten, mit der aus unterschiedlichen Perspektiven eine wissenschaftliche Arbeit betrachtet werden kann. Erst die Verknüpfung beider Perspektiven lässt ein umfassendes Gesamtbild entstehen, durch das *belastbare* Aussagen über die Qualität einer wissenschaftlichen Arbeit ermöglicht werden.

Literaturverzeichnis

- Abate, T. (1995). What's the verdict on peer review? *Ethics in Research*, 1, 1.
Adam, D. (2002). The counting house. *Nature*, 415(6873), 726-729.

- Barbui, C., Cipriani, A., Malvini, L. & Tansella, M. (2006). Validity of the impact factor of journals as a measure of randomized controlled trial quality. *Journal of Clinical Psychiatry*, 67(1), 37-40.
- Berghmans, T., Meert, A. P., Mascaux, C., Paesmans, M., Lafitte, J. J. & Sculier, J. P. (2003). Citation indexes do not reflect methodological quality in lung cancer randomised trials. *Annals of Oncology*, 14(5), 715-721.
- Biagioli, M. (2002). From book censorship to academic peer review. *Emergences*, 12(1), 11-45.
- Bozeman, B., Dietz, J. S. & Gaughan, M. (2001). Scientific and technical human capital: an alternative model for research evaluation. *International Journal of Technology Management*, 22(7-8), 716-740.
- Butler, L. (2004). What happens when funding is linked to publication counts? In H. F. Moed, W. Glänzel & U. Schmoch (Hrsg.), *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems* (S. 389-405). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Campanario, J. M. & Acedo, E. (2007). Rejecting highly cited papers: the views of scientists who encounter resistance to their discoveries from other scientists. *Journal of the American Society for Information Science and Technology*, 58(5), 734-743.
- Daniel, H.-D. (2001). *Wissenschaftsevaluation. Neuere Entwicklungen und heutiger Stand der Forschungs- und Hochschulevaluation in ausgewählten Ländern (CEST 2001/2)*. Bern, Switzerland: Center for Science and Technology Studies (CEST).
- Daniel, H.-D. (2006). Pro Peer Review: 5 Thesen. In Hochschulrektorenkonferenz Projekt Qualitätssicherung (Hrsg.), *Von der Qualitätssicherung der Lehre zur Qualitätsentwicklung als Prinzip der Hochschulsteuerung* (Band 1, S. 185-192). Bonn, Germany: Hochschulrektorenkonferenz, Projekt Qualitätssicherung.
- de Solla Price, D. (1965). Networks of scientific papers: the pattern of bibliographic references indicates the nature of the scientific research front. *Science*, 149(3683), 510-515.
- Döring, T. F. (2007). Quality evaluation needs some better quality tools. *Nature*, 445(7129), 709.
- 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.
- European Science Foundation. (2007). *Peer review. Its present and future state (Conference report)*. Strasbourg, France: European Science Foundation (ESF).
- Finn, C. E. (2002). The limits of peer review. *Education Week*, 21(34), 30-34.
- Geisler, E. (2000). *The metrics of science and technology*. Westport, CT, USA: Quorum Books.
- Gläser, J. (2006). *Wissenschaftliche Produktionsgemeinschaften. Die soziale Ordnung der Forschung*. Frankfurt am Main, Germany: Campus.
- Graefen, G. & Thielmann, W. (2007). Der wissenschaftliche Artikel. In P. Auer & H. Baßler (Hrsg.), *Reden und Schreiben in der Wissenschaft* (S. 68-97). Frankfurt am Main, Germany: Campus.
- Grivell, L. (2006). Through a glass darkly - the present and the future of editorial peer review. *EMBO Reports*, 7(6), 567-570.
- Hackett, E. J. & Chubin, D. E. (2003). Peer review for the 21st century: applications to education research. Paper presented at the Peer review of education research grant applications. Implications, considerations, and future directions, Washington, DC, USA.
- Hanlin, L. (1984). *Ausdifferenzierung der Wissenschaftsbewertung in der Wissensproduktion*. München, Germany: Minerva.
- Hansson, F. (2002). How to evaluate and select new scientific knowledge? Taking the social dimension seriously in the evaluation of research quality. *VEST*, 15(2-3), 27-52.
- Hornbostel, S. (1997). *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*. Opladen, Germany: Westdeutscher Verlag.
- Horrobin, D. F. (2001). Something rotten at the core of science? *Trends in Pharmacological Sciences*, 22(2), 51-52.

- Jones, P. & Sizer, J. (1990). The universities funding council's 1989 research selectivity exercise. *Beiträge zur Hochschulforschung*(4), 309-348.
- Kornhuber, H. H. (1988). Mehr Forschungseffizienz durch objektivere Beurteilung von Forschungsleistungen. In H.-D. Daniel & R. Fisch (Hrsg.), *Evaluation von Forschung. Methoden - Ergebnisse - Stellungnahmen* (S. 361-382). Konstanz, Germany: Universitätsverlag Konstanz.
- Marsh, H. W. & Ball, S. (1991). Reflections on the peer review process. *Behavioral and Brain Sciences*, 14(1), 157-158.
- 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, v. 93)*. Philadelphia, PA, USA: American Philosophical Society.
- Meho, L. I. (2007). *The rise and rise of citation analysis*. Retrieved July 16, 2007, from <http://arxiv.org/abs/physics/0701012>
- Mulligan, A. & Mabe, M. (2006, April, 3-5). Journal futures: researcher behaviour at early internet maturity. Paper presented at the UK Serials Group (UKSG) Annual Conference, University of Warwick, UK.
- Neidhardt, F. (2006). Fehlerquellen und Fehlerkontrollen in den Begutachtungssystemen der Wissenschaft. *iFQ-Working paper*(1), 7-13.
- Neuhaus, C. & Daniel, H.-D. (2008). Data sources for performing citation analysis - an overview. *Journal of Documentation*, 64(2), 193-210.
- Research Evaluation and Policy Project. (2005). *Quantitative indicators for research assessment – a literature review (REPP discussion paper 05/1)*. Canberra, Australia: 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, MA, USA: The Ross Company.
- Roy, R. (1985). Funding science - the real defects of peer-review and an alternative to it. *Science Technology & Human Values*, 52, 73-81.
- Schneider, W. L. (1988). Grenzen der Standardisierbarkeit bei der Bewertung von Forschungsergebnissen. In H.-D. Daniel & R. Firsch (Hrsg.), *Evaluation von Forschung. Methoden - Ergebnisse - Stellungnahmen* (S. 433-447). Konstanz, Germany: Universitätsverlag Konstanz.
- Shadbolt, N., Brody, T., Carr, L. & Harnad, S. (2006). The Open Research Web: a preview of the optimal and the inevitable. In N. Jacobs (Hrsg.), *Open Access: key strategic, technical and economic aspects* (S. 195-208). Oxford, UK: Chandos.
- Shapiro, B. J. (2002). Testimony in seventeenth-century English natural philosophy: legal origins and early development. *Studies in History and Philosophy of Science*, 33A(2), 243-263.
- Shashok, K. (2005). Standardization vs diversity: how can we push peer review research forward? *Medscape General Medicine*, 7(1), 11.
- Smith, L. C. (1981). Citation analysis. *Library Trends*, 30(1), 83-106.
- Smith, R. (1988). Problems with peer review and alternatives. *British Medical Journal*, 296(6624), 774-777.
- Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99(4), 178-182.
- Steele, C., Butler, L. & Kingsley, D. (2006). The Publishing imperative: the pervasive influence of publication metrics. *Learned Publishing*, 19(4), 277-290.
- van Raan, A. F. J. (1999). Advanced bibliometric methods for the evaluation of universities. *Scientometrics*, 45(3), 417-423.
- van Raan, A. F. J. (2004). Measuring science. Capita selecta of current main issues. In H. F. Moed, W. Glänzel & U. Schmoch (Hrsg.), *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems* (S. 19-50). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Weingart, P. (2001). *Die Stunde der Wahrheit? Zum Verhältnis der Wissenschaft zu Politik, Wirtschaft und Medien in der Wissensgesellschaft*. Weilerswist, Germany: Velbrück Wissenschaft.

- Weingart, P. (2005a). Das Ritual der Evaluierung und die Verführbarkeit. In P. Weingart (Hrsg.), *Die Wissenschaft der Öffentlichkeit: Essays zum Verhältnis von Wissenschaft, Medien und Öffentlichkeit* (S. 102-122). Weilerswist, Germany: Velbrück.
- Weingart, P. (2005b). Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics*, 62(1), 117-131.
- Wright, M. & Armstrong, J. S. (2007). *Verification of citations: faulty towers of knowledge?* Retrieved August 6, from <http://mpira.ub.uni-muenchen.de/4149/>
- Young, S. N. (2003). Peer review of manuscripts: theory and practice. *Journal of Psychiatry & Neuroscience*, 28(5), 327-330.
- Ziman, J. (2000). *Real science. What it is, and what it means*. Cambridge, UK: Cambridge University Press.