Peer Review

DOI: 10.1002/anie.200800513

The Effectiveness of the Peer Review Process: Inter-Referee Agreement and Predictive Validity of Manuscript Refereeing at *Angewandte Chemie*

Lutz Bornmann and Hans-Dieter Daniel**

manuscript refereeing · peer review · predictive validity

1. Introduction

In modern science, peer review has advanced to become the most important instrument for assessing scientific work.[1] Supporters of the peer review process deem it indispensable, because only active researchers within the same research field—peers—are in a position to assess the scientific quality of their colleagues' research.[2] Critics of peer review see as weaknesses of the process that different referees' assessments of one and the same piece of scholarly work rarely agree (problem of inter-referee agreement), and that recommendations and decisions in peer review show systematic biased judgments, resulting in little correlation between the decision to publish and the impact of the refereed manuscript following publication (problem of predictive validity).[3,4]

Research on peer review, which in recent years has examined the criticisms of the peer review process, deals for the most part with journal peer review[5–8] and somewhat less frequently with peer review for research grant proposals and fellowship applications.[9–11] For Godlee and Jefferson, "the biggest surprise is how little we know about its workings. For a system that demands ever increasing rigor and levels of proof from scientists, it remains itself remarkably untouched by the rigors of science."[12] Many peer review studies have methodological weaknesses, and "most of the publications on journal peer review are more opinion than research, often the ruminations of a former editor. Likewise, most of the many letters to editors on the topic, the comments of one kind or another are predominantly opinion."[13]

The present study examines whether the criticisms of the peer review process are justified at a journal that publishes original research papers. In the study *Guardians of Science*,[14] Daniel evaluated the peer review process at *Angewandte Chemie* in the mid-1980s. We once again examined the quality of peer review at *Angewandte Chemie* according to the criteria of reliability (agreement among referees) and predictive validity, using an optimized study design and modern methods of statistical data analysis.

To investigate reliability we determined the degree of agreement among referees.[15] Judgments are called reliable if there is agreement of the publication recommendations of several independent referees for a manuscript.[16]

To check the predictive validity of the peer review process, studies on manuscript review use, for lack of other operationalizable indicators, the citation counts of manuscripts accepted for publication and manuscripts rejected by a journal but then published elsewhere.[17] According to van Raan, citations provide "a good to even very good quantitative impression of at least one important aspect of quality, namely international impact."[18] For Lindsey, citations are "our most *reliable* convenient measure of quality in science—a measure that will continue to be widely used."[19] Scientific judgments of whether journal manuscripts are worthy of publication are said to show predictive validity if there is a statistically significant difference between the citation rates of manuscripts accepted for publication and those of manuscripts rejected by a journal but then published elsewhere.[14]

2. Methods

Angewandte Chemie is a journal of the German Chemical Society (Gesellschaft Deutscher Chemiker (GDCh), Frankfurt am Main) and is published by Wiley-VCH (Weinheim). A Communication (*Zuschrift*) submitted to the journal normally undergoes internal and external refereeing. First, a journal editor evaluates the importance of the Communication for the development of a research area within chemistry (internal refereeing). If in the opinion of the editor it is very important, the Communication is sent to several (usually three)[20] independent external referees, who are requested to make a recommendation on an evaluation form as to whether the Communication should be published (one of the questions on the form is, "Do you recommend acceptance of the Communication?") and to explain their reasons for the recommen-

[*] Dr. L. Bornmann, Prof. Dr. H.-D. Daniel
ETH Zurich
Professorship for Social Psychology and Research on Higher Education
Zähringerstr. 24, 8092 Zurich (Switzerland)
Fax: (+41) 44-632-12-83
E-mail: bornmann@gess.ethz.ch
Prof. Dr. H.-D. Daniel
University of Zurich, Director of the Evaluation Office
Mühlegasse 21, 8001 Zurich (Switzerland)
Fax: (+41) 44-634-43-79
E-mail: daniel@evaluation.uzh.ch

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200800513.

dation on a sheet provided for comments. Based on the referees' recommendations, the journal editor decides whether to accept or reject a Communication for publication.[21]

For the investigation of peer review at *Angewandte Chemie* we used information on a total of 1899 Communications that were refereed in the year 2000. The information was taken from archived material that was stored electronically by the publisher. By using the Communications refereed in 2000, we selected a somewhat older manuscript cohort as the database, since for an examination of predictive validity there should be a time interval of several years between reviewing of a manuscript and measurement of the indicator for scientific quality (in this case: citations).[22] In addition to the internal review by the journal editor, a total of 4593 external reviews on the 1899 Communications were made using evaluation forms or comments sheets. On average, therefore, there were 2.4 external reviews available for each Communication. Based on these recommendations, 46% ($n=878$) of the 1899 Communications were accepted for publication in *Angewandte Chemie*, and 54% ($n=1021$) were rejected.

Table 1: Agreement between referees in their responses to the question: "Do you recommend acceptance of the Communication?"[a,b]

Number of referees per Communication	Number of Communications reviewed in common	Observed agreement [%]	Expected agreement [%]	κ[c]	Confidence interval[d]
Two referees	952	41.8	31.8	0.15^{u}[e]	0.10–0.19
	952	69.1	61.2	0.21^{w}[e]	0.16–0.25
Three referees	535			0.10^{u}[e]	0.07–0.14
From two to five referees	1507			0.12^{u}[f]	0.09–0.15

[a] Response categories: "Yes, without alterations", "Yes, after minor alterations", "Yes, but only after major alterations", and "No". [b] Of 1899 Communications, 1507 could be included in the statistical analysis of agreement between *Angewandte Chemie* referees: Of 4593 external reviews, 3023 are available with an evaluation form containing the question, "Do you recommend acceptance of the Communication?," to which the referees responded "Yes, without alterations", "Yes, after minor alterations", "Yes, but only after major alterations", or "No". These 3023 reviews pertain to 1840 Communications. Since testing for inter-referee agreement requires at least two reviews per Communication, 952 Communications with two reviews, 535 Communications with three reviews, 19 Communications with four reviews, and one Communication with five reviews could be included in the analysis (for 333 Communications, only one review is available). In addition to the kappa coefficient for Communications with two to five reviews ($n=1507$), coefficients were calculated for Communications with two ($n=952$) and three ($n=535$) reviews. For Communications for which four or five reviews are available, these coefficients were not calculated, as for 19 Communications (four reviews) and one Communication (five reviews) the number of cases is too small. The table shows weighted and unweighted κ values. In contrast to unweighted κ, the weighted κ value additionally takes into account that where there is a lack of agreement between the responses of two referees, there can be different degrees of disagreement. In the analysis, a weight of 0.6667 was assigned to those Communications where the referees show "two-thirds agreement" (that is, the referees chose nearby response categories, such as "Yes, without alterations" and "Yes, after minor alterations"). A weight of 0.3333 was assigned in the case of one-third agreement (for example, "Yes, without alterations" and "Yes, after major alterations"). In the analysis, a weight of 0 (that is, no weight) was used when the referees' responses were completely contrary and a weight of 1 was used when the referees' responses agreed completely. [c] u = unweighted; w = weighted. [d] Based on 1000 sample tables. [e] $p<0.05$. [f] Significance tests could not be calculated.

3. Results

3.1. Inter-Referee Reliability of the Peer Review Process

3.1.1. Agreement among Referees' Recommendations

Table 1 shows kappa coefficients describing agreement between the responses of two or more referees per Communication to the question, "Do you recommend acceptance of the Communication?" (response categories: "Yes, without alterations", "Yes, after minor alterations", "Yes, but only after major alterations", and "No"). Kappa (κ) is a statistical measure of level of agreement between two or more raters.[23] If the raters are in complete agreement, then $\kappa=1$; if κ is near 0, the observed level of agreement is not much higher than would be expected by chance: "Multiplied by 100, κ indicates the percentage by which two raters' agreement exceeds the agreement that could be expected from chance."[23] As Table 1 shows, for peer review at *Angewandte Chemie* the kappa coefficients range from 0.10 to 0.21, depending on the κ statistic (weighted or unweighted) and the number of referees (between two and five) that made a recommendation on a Communication. In other words, the referees show agreement in their responses on 10–21% more manuscripts than would have been expected by chance. Although different κ statistics were calculated and the responses of a differing number of referees per Communication were included in the analysis, the resulting values hardly differed.[24] According to guidelines for the interpretation of κ,[23] the coefficients in Table 1 indicate a low level of agreement among referees' recommendations. This result confirms Daniel's[25] findings on peer review at *Angewandte Chemie* in the mid 1980s. The other studies that are available on inter-referee reliability of publication recommendations at other journals also report mainly low kappa coefficients.[8]

3.1.2. Agreement between Referees' Recommendations and Editorial Decisions

In the manuscript peer review process, the journal editor makes the decision to accept or reject a Communication for publication on the basis of the referees' recommendations.[26] If the recommendations of the referees on a Communication agree, the editor can make a decision in accord with both referees. But what is the editor's decision when the referees' recommendations on a manuscript differ (such as, "Yes, after minor alterations" and "No")? To answer this question, we produced a configuration of the referees' responses for each Communication to the question, "Do you recommend acceptance of the Communication?"

Table 2: Acceptance and rejection of a Communication by response configuration of two referees to the question: "Do you recommend acceptance of the Communication?"[a,b]

Response configuration of two referees	Number of Communications	Acceptance ($n=302$)	Rejection ($n=416$)
"Yes, without alterations" – "Yes, without alterations"	15	**$^{+}$100.0**	**$^{-}$.0**
"Yes, without alterations" – "Yes, after minor alterations"	96	**$^{+}$100.0**	**$^{-}$.0**
"Yes, after minor alterations" – "Yes, after minor alterations"	100	**$^{+}$98.0**	**$^{-}$2.0**
"Yes, after minor alterations" – "Yes, but only after major alterations"	44	**$^{+}$97.7**	**$^{-}$2.3**
"Yes, without alterations" – "Yes, but only after major alterations"	18	**$^{+}$94.4**	**$^{-}$5.6**
"Yes, but only after major alterations" – "Yes, but only after major alterations"	9	44.4	55.6
"Yes, without alterations" – "No"	24	37.5	62.5
"Yes, after minor alterations" – "No"	73	**$^{-}$24.7**	75.3
"Yes, but only after major alterations" – "No"	91	**$^{-}$2.2**	**$^{+}$97.8**
"No" – "No"	248	**$^{-}$.0**	**$^{+}$100.0**
Total	718	42.1	57.9

[a] Response categories: "Yes, without alterations", "Yes, after minor alterations", "Yes, but only after major alterations", and "No" (row percentage, sorted in decreasing order by percentages in the "Acceptance" column). [b] Of 1899 Communications, 718 could be included in the analysis. To test in the statistical analysis whether the editor's decision is based on the recommendations of the referees, it is necessary that for all of the recommendations pertaining to a Communication, the referees responded to the question, "Do you recommend acceptance of the Communication?" In contrast to Table 1, here no manuscripts were included in the statistical analysis where a court-of-appeal referee reviewed the manuscript, and no case was included where the review opinion pertained to a revised manuscript or an appeal that was submitted by an author in response to rejection of a manuscript. Of the 1069 Communications for which these conditions are fulfilled, there are two reviews for 718 Communications, three reviews for 241 Communications, and four reviews for seven Communications, (only one review each is available for 103 Communications). Since a statistical analysis that includes a response configuration with three ($n=241$) and four ($n=7$) reviews per Communication is not advisable because of the small number of cases and the large number of possible configurations, the analysis was carried out on 718 Communications, for each of which there were two reviews. For this subgroup, agreement between the referees' recommendations, with kappa coefficients of 0.27 (unweighted κ) and 0.43 (weighted κ), is stronger than in the whole group (see Table 1). With regard to the response configurations, the difference between decisions to accept or reject is statistically significant; χ^2 test: χ^2 (9, $n=718$) $=606.2$, $p<0.0001$. Cells with standardized residuals greater than 2 (or less than -2) are shown in the table in bold type. Residuals are a measure of how strongly the observed frequencies deviate from the expected frequencies.

(for example, Referee 1: "Yes, without alterations" and Referee 2: "Yes, after minor alterations") and examined the correlation between this response configuration and the editor's decision.

As the results in Table 2 show, response configurations in which the referees recommend accepting a Communication for publication (with no, minor, or major alterations) are associated with the editorial decision to accept for publication; for the rejected Communications it is the other way around. Therefore, with regard to the referees' response configurations, the test statistics indicate a highly significant difference between the decisions to accept and reject (see Table 2). There is thus statistical evidence that a Communication is as a rule accepted by the journal editor when both referees (Referee 1 and Referee 2) recommend accepting it for publication.

3.2. Predictive Validity of the Peer Review Process

The Supporting Information for this Essay provides extensive information on which journals Communications rejected for publication by *Angewandte Chemie* were later published in and on how greatly the authors altered the Communications for publication in other journals. Here we examine the predictive validity of the peer review process at *Angewandte Chemie*, on the basis of the mean citation counts of manuscripts accepted for publication and manuscripts rejected by the journal but then published elsewhere and on the Journal Impact Factors (JIFs) of the journals in which the rejected Communications were later published. JIFs are published by Thomson Reuters (Philadelphia, PA, USA) in the Journal Citation Reports (JCR) and are a measure of the "average" response of the scientific community to an article in a journal.[27] Published annually by Thomson Reuters, the JIF is the quotient of citations and number of citable items: the JIF is determined for articles published in a journal in a two-year period and their citations indexed in the year thereafter. The number of times that the articles were cited is then divided by the number of citable items (for *Angewandte Chemie*, these are essentially the Communications and a small number of Review articles) published in the previous two years.[28]

According to the results of investigation in two literature databases, of 1021 Communications rejected for publication by *Angewandte Chemie*, 959 were published as contributions (93.9%) in other journals, seven as patents, and two as contributions to anthologies. No publication information was found for 53 (5.2%) of the rejected Communications. The 959 Communications that later appeared in other journals were published in 136 journals. Fifty or more rejected Communications each were published in the journals *Chemical Communications* ($n=119$), *Organic Letters* ($n=91$), *Journal of the American Chemical Society* ($n=70$), *Tetrahedron Letters* ($n=60$), and *Organometallics* ($n=50$). No alterations or only minor alterations were made to approximately three-quarters of the rejected Communications for publication elsewhere. The Communications rejected for publication by *Angewandte Chemie* were published in other journals within a time period of seven years (that is, between 2000 and 2006).

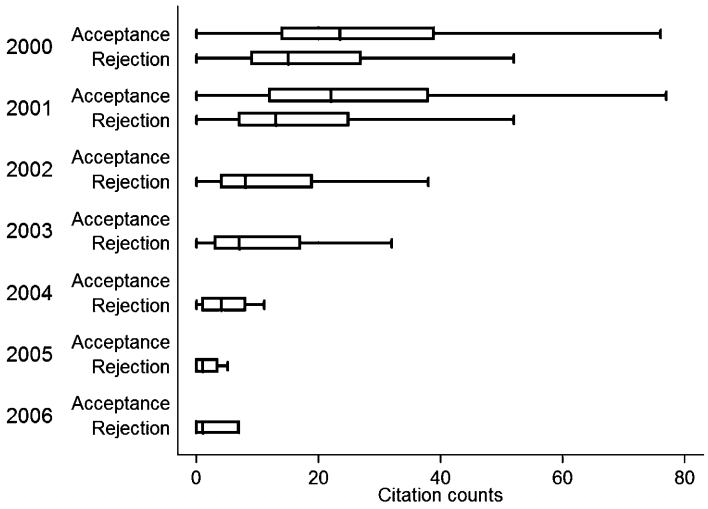

Figure 1. Box plots: Mean citation counts (the vertical line within each box indicates the median) by publication year of Communications that were accepted and Communications that were rejected but published elsewhere, and by the editor's decision (outliers not shown). Of 1899 Communications, 1827 could be included in the analysis: 62 of the Communications rejected by *Angewandte Chemie* were—according to the results of our search—not published elsewhere. For 10 of the rejected Communications that were published elsewhere, no citations could be found in Scopus. The difference between the mean citation counts for accepted (median = 23.5) and rejected (median = 15) Communications that were published in the year 2000 is statistically significant; Mann–Whitney U-test: $Z(n=778) = -6.29$, $p < 0.0001$. The difference between the mean citation counts for accepted (median = 22) and rejected (median = 13) Communications that were published in 2001 is also statistically significant; Mann–Whitney U-test: $Z(n=885) = -7.45$, $p < 0.0001$. After 2001 there were only Communications that were rejected but published elsewhere.

According to the JCR for 2006, the JIFs of these other journals ranged from less than 1 (for example, *CHIMIA* and *Chinese Chemical Letters*) to 9.96 (*Nano Letters*). As measured by these JIFs, the *Angewandte Chemie* editorial decisions to accept or reject are highly valid. None of the total of 956 rejected Communications was published in a journal with a higher JIF than *Angewandte Chemie* (JIF 2006 = 10.23);[*] all of the rejected Communications were published in a journal with a lower JIF (with the exception of 16 rejected Communications that were accepted for publication by *Angewandte Chemie* after resubmission). This same result was found by Daniel[14] for Communications rejected by *Angewandte Chemie* in 1984. These findings confirm Cronin and McKenzie's general observation that manuscripts that are rejected by prestigious journals with high JIFs are usually later submitted to (and published by) journals with lower JIFs: "It is widely recognized that there is an informal journal pecking order in almost every discipline, and that a manuscript rejected by a high-ranking journal will often be resubmitted to one of lesser repute."[29]

However, the JIF is only a very rough measure for determining predictive validity, because all of the contributions in a journal are characterized by an average value.[30] It thus underestimates the citations of the top cited articles while exaggerating the number of citations of articles that are not or infrequently cited.[14] For this reason, going beyond JIFs, we determined how frequently the manuscripts accepted for publication and the manuscripts rejected but later published elsewhere were cited after being published up to the year 2006. The citation counts for the individual manuscripts were investigated in Scopus, a research literature database.[31] Scopus, which is a new multidisciplinary database provided by Elsevier (Amsterdam, The Netherlands), contains over 33 million abstracts and references from 15000 peer-reviewed journals from more than 4000 international publishers and is currently the largest multidisciplinary database of research.[32] On the basis of results by Norris and Oppenheim,[33] Scopus can be considered an equally matched alternative to the Web of Science (WoS) from Thomson Reuters.

As the results for publication years 2000 and 2001 in Figure 1 show, the Communications accepted for publication were clearly more frequently cited than the rejected Communications (an analysis of the citations in the Science Citation index (SCI, Thomson Reuters) and Chemical Abstracts (Chemical Abstracts Services, Columbus, OH, USA) yielded similar results).[34] The comparison could not be conducted for publication years 2002 to 2006, as after 2001 there are only manuscripts that had been rejected from *Angewandte Chemie* but published elsewhere, but no manuscripts accepted for publication (reviewed in the year 2000) (the period of time to publication at *Angewandte Chemie* is comparatively short). The differences in the means between the citation counts are statistically significant (see Figure 1). This result corresponds with the finding that the rejected manuscripts were later published in journals with a lower JIF (see above).

[*] Based on JIFs published by Thomson Reuters in the JCRs for the years 2000 and 2001 (that is, for the years in which most of the rejected manuscripts were published elsewhere), only one rejected manuscript was published in a journal having a higher JIF, namely, in *Proceedings of the National Academy of Sciences of the United States of America* (*PNAS*) (JIF 2000 = 10.79; JIF 2001 = 10.90). The JIF of *Angewandte Chemie* was 8.55 for 2000 and 8.26 for 2001. In the JCR for 2006, *Angewandte Chemie* (JIF 2006 = 10.23) had a higher JIF than *PNAS* (JIF 2006 = 9.64).

Since—as the citation analysis results make clear—the citation counts of manuscripts accepted and rejected for publication differ statistically significantly, the findings of the citation analysis confirm the predictive validity of the peer review process at *Angewandte Chemie*.

4. Discussion

Against the background of criticisms of the peer review process, the present study investigated peer review at *Angewandte Chemie* according to the criteria of inter-referee agreement and predictive validity. With regard to the reliability of the peer review process, the kappa coefficients (between 0.10 and 0.21) indicated a low level of agreement between the referees' recommendations concerning acceptance or rejection of a Communication. Similar findings on peer review at *Angewandte Chemie* were reported by Daniel in an earlier study.[25]

Although a high level of agreement among assessments is generally seen as desirable, when it comes to peer review some researchers, such as Bailar, view agreement as detrimental to the review process: "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."[35] Although selecting referees according to the principle of complementarity (for example, choosing a generalist and a specialist) will lower inter-referee agreement, the validity of the process can gain, according to Langfeldt: "Low inter-reviewer agreement on a peer panel is no indication of low validity or low legitimacy of the assessments. In fact, it may indicate that the panel is highly competent because it represents a wide sample of the various views on what is good and valuable research."[36] Differing recommendations in manuscript refereeing are not necessarily a sign of disagreement and can be due to the differing paradigmatic positions ("schools"), approaches, and mentalities of the referees.[37] In addition, referees can tend to be more critical or more lenient in their judgments;[38] they direct their attention, writes Eckberg, to "different points, and may draw different conclusions about 'worth'".[39]

The examination of the fate of Communications that were rejected for publication by *Angewandte Chemie* showed that, with a very few exceptions, all of the Communications were subsequently published elsewhere. They were always published in journals that have a lower JIF than *Angewandte Chemie*. The citation analysis showed that the Communications accepted by *Angewandte Chemie* were on average cited statistically more frequently than manuscripts rejected by the journal and published elsewhere. Both of these findings attest to the predictive validity of the peer review process at *Angewandte Chemie*. Similar findings have been reported to date not only by Daniel[14] for *Angewandte Chemie* but also by Lock,[40] Wilson,[41] and Opthof, Furstner, van Geer, and Coronel[42] for the *British Medical Journal*, the *Journal of Clinical Investigation*, and *Cardiovascular Research*.

Although the surrounding conditions of the peer review process at *Angewandte Chemie* have definitely changed since the mid-1980s (that is, since the study by Daniel)[14] (for example, dramatic increase in the number of Communications submitted and clear decrease in the acceptance rate), the results of the present investigation of inter-referee agreement and predictive validity demonstrate once again the high quality of the peer review process at *Angewandte Chemie*.

We would like to thank Dr. Christophe Weymuth (formerly at the Institute of Organic Chemistry at the University of Zurich, now at BIOSYNTH AG in Switzerland) for the investigation of the manuscripts rejected by Angewandte Chemie and published elsewhere. We also thank Dr. Andras Schubert (Hungarian Academy of Sciences) for providing the Journal Impact Factors published by Thomson Reuters in the Journal Citation Reports for the years 2000 and 2001 for Angewandte Chemie and for the journals in which manuscripts were published that had been rejected by Angewandte Chemie and published elsewhere. We thank Dr. Peter Gölitz, Editor-in-Chief of Angewandte Chemie, the Editorial Board of Angewandte Chemie, and the German Chemical Society (GDCh) for permission to conduct the evaluation of the peer review process of the journal, and to the members of the editorial office for their generous support during the study. The entire study, which is also investigating quality assurance at open-access journals, is supported by the Max Planck Society. The authors would like to thank three anonymous referees for their valuable comments.

Received: May 2, 2008
Revised: March 27, 2008
Published online: July 25, 2008

[1] J. Ziman, *Real Science. What it is, and what it means*, Cambridge University Press, Cambridge, **2000**.
[2] M. Eisenhart, *Res. Sci. Educ.* **2002**, *32*, 241.
[3] P. F. Ross, *The Sciences' Self-management: Manuscript Refereeing, Peer Review, and Goals in Science*, The Ross Company, Todd Pond, **1980**.
[4] H. J. Eysenck, S. B. G. Eysenck, *Pers. Indiv. Differ.* **1992**, *13*, 393.
[5] J. M. Campanario, *Sci. Commun.* **1998**, *19*, 277.
[6] J. M. Campanario, *Sci. Commun.* **1998**, *19*, 181.
[7] J. Overbeke, E. Wager in *Peer Review in Health Sciences*, 2nd ed. (Eds.: F. Godlee, T. Jefferson), BMJ, London, **2003**, p. 45.
[8] A. C. Weller, *Editorial Peer Review: Its Strengths and Weaknesses*, Information Today, Medford, **2002**.
[9] V. Demicheli, C. Pietrantonj in *The Cochrane Library*, *Issue 1*, Wiley, Chichester, **2004**.
[10] S. Wessely, *Lancet* **1998**, *352*, 301.
[11] L. Bornmann, H.-D. Daniel in *Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung* (Eds.: S. Schwarz, U. Teichler), Campus, Frankfurt, **2003**, p. 211.
[12] *Peer Review in Health Sciences*, 2nd ed. (Eds.: F. Godlee, J. Jefferson), BMJ, London, **2003**.
[13] M. F. Stieg Dalton, *ARIST* **1995**, *30*, 213.
[14] H.-D. Daniel, *Guardians of Science. Fairness and Reliability of Peer Review*, Wiley-VCH, Weinheim, **1993/2004**. Published online 16 July 2004, Wiley Interscience, DOI: 10.1002/3527602208.
[15] D. V. Cicchetti, *Behav. Brain Sci.* **1991**, *14*, 119.
[16] J. Ziman, *Reliable Knowledge: An Exploration of the Grounds for Belief in Science*, Cambridge University Press, Cambridge, **1991**.
[17] H.-D. Daniel, *Learn. Publ.* **2005**, *18*, 143.

[18] A. F. J. van Raan, *Scientometrics* **1996**, *36*, 397.
[19] D. Lindsey, *Scientometrics* **1989**, *15*, 189.
[20] Anonymous, *Chem. Eng. News Peer Review* **2008**, *86*(6), 48.
[21] A. M. Coghill, L. R. Garson, *The ACS Style Guide*, 3rd ed., American Chemical Society, Washington, **2006**.
[22] H. P. F. Peters, A. F. J. van Raan, *J. Am. Soc. Inf. Sci.* **1994**, *45*, 39.
[23] A. von Eye, E. Y. Mun, *Analyzing Rater Agreement. Manifest Variable Methods*, Lawrence Erlbaum, Mahwah, **2005**.
[24] U. W. Jayasinghe, H. W. Marsh, N. Bond, *Educ. Eval. Policy. Anal.* **2001**, *23*, 343.
[25] H.-D. Daniel, *Angew. Chem.* **1993**, *105*, 247; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 234.
[26] V. Bakanic, C. McPhail, R. J. Simon, *Am. Sociol. Rev.* **1987**, *52*, 631.
[27] L. Bornmann, L. Leydesdorff, W. Marx, *CHIMIA* **2007**, *61*, 104.
[28] W. Marx, H. Schier, *Phys. Bl.* **2001**, *57*, 25.
[29] B. Cronin, G. McKenzie, *J. Doc.* **1992**, *48*, 310.
[30] T. Braun, I. Dióspatonyi, S. Zsindely, E. Zádor, *Scientometrics* **2007**, *71*, 541.
[31] J. Bar-Ilan, M. Levene, A. Lin, *J. Informetrics* **2007**, *1*, 26.
[32] F. de Moya-Anegón, Z. Chinchilla-Rodríguez, B. Vargas-Quesada, E. Corera-Álvarez, F. Muñoz-Fernández, A. González-Molina, V. Herrero-Solana, *Scientometrics* **2007**, *73*, 53.
[33] M. Norris, C. Oppenheim, *J. Informetrics* **2007**, *1*, 161.
[34] L. Bornmann, H.-D. Daniel, *J. Am. Soc. Inf. Sci. Technol.* **2008**, *59*, 830.
[35] J. C. Bailar, *Behav. Brain Sci.* **1991**, *14*, 137.
[36] L. Langfeldt, *Soc. Stud. Sci.* **2001**, *31*, 820.
[37] R. N. Kostoff, *Scientometrics* **1995**, *34*, 163.
[38] S. S. Siegelman, *Radiology* **1991**, *178*, 637.
[39] D. L. Eckberg, *Behav. Brain Sci.* **1991**, *14*, 145.
[40] S. Lock, *A Difficult Balance: Editorial Peer Review in Medicine*, ISI, Philadelphia, **1985**.
[41] J. D. Wilson, *J. Clin. Invest.* **1978**, *61*, 1697.
[42] T. Opthof, F. Furstner, M. van Geer, R. Coronel, *Cardiovasc. Res.* **2000**, *45*, 255.

