Learned Publishing, 22, 117–125
doi:10.1087/2009207

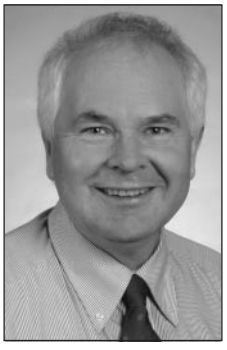
RESEARCH ARTICLE

The luck of the referee draw: the effect of exchanging reviews

Lutz BORNMANN and Hans-Dieter DANIEL

ABSTRACT. In journal peer review, editorial decisions on submitted manuscripts are informed by referees' expert recommendations; however, the choice of referees may affect these decisions. Using data from Angewandte Chemie International Edition *(AC-IE), this study tested what would have happened if referee reports had been received in a different order. In* AC-IE*'s peer-review process, a manuscript is generally published only if two referees rate the results of the study as important and also recommend publication in the journal (what we have called the 'clear-cut' rule). For 23% of those manuscripts for which a third referee 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.*

Lutz Bornmann



Hans-Dieter Daniel

Introduction

In academic science today, peer review of contributions to the primary research literature is the principal social mechanism for quality control.[1] Since peer review is so central to what is published and where, and since so much hinges on peer review in and outside of science, it is essential that it is carried out well and professionally.[2]

The authors carried out an extensive research project to establish whether the common criticisms of the peer-review process were justified in the case of *Angewandte Chemie International Edition* (*AC-IE*), a journal that publishes original research papers. In the context of this project, archive data were collected and evaluated on 1,899 submitted Communications that were reviewed in the year 2000.[3,4] *AC-IE* is a journal of the German Chemical Society (Gesellschaft Deutscher Chemiker), Frankfurt am Main, Germany, and is published by Wiley-VCH (Weinheim, Germany). Peer review was introduced at *AC-IE* in 1982, primarily in conjunction with Communications (called 'letters' or 'notes' in other journals). In the *AC-IE* peer-review process, for most submissions a manuscript is published only if two external referees rate the results of the study reported in the manuscript as (very) important and also recommend publication in the journal – what we have called the 'clear-cut' rule.[5]

Against the background of this 'clear-cut' rule, this study investigates the question of what would have happened if an editorial decision on a submission had been based on a third referee report in place of either the first or the second. Even though the 'clear-cut' rule is based on two referee reports, submitted Communications generally go out to three referees in total.[6] The reason given for this is that, in the experience of the *AC-IE* editors, in today's increasingly busy climate

there are many referees who are unable to review papers because of other commitments. Also, the editors have a responsibility to authors to make a rapid, fair decision on the outcome of manuscripts. In addition to 385 submissions where three referees' reports were available at the decision time, in our database there are also 270 submissions on which the editorial decision was made based on two referee reports, and for which a third referee report arrived too late at the editorial office to be considered in the editorial decision.

Methods

Communications at AC-IE

Communications are short reports (limited to six manuscript pages) dealing with work in progress or recently concluded experimental or theoretical investigations in any of the various branches of chemistry. A Communication is expected to be of broad general interest, due to its significance, novelty, or wide applicability, or at least of special utility in the development of some important areas of research. It must also be written in such a way that even a non-specialist will recognize the significance that the author attaches to the findings.[7] Scientists publish research results in the form of Communications rapidly to establish priority claims[8] for their findings. The *Journal of the American Chemical Society*[9] stresses this criterion of urgency in its own definition: 'Communications are restricted to reports of unusual urgency, timeliness, significance, and broad interest.' The immediacy and the short length of the reports of research findings are hence important features of Communications, distinguishing them from research articles and reviews.[10] From 1984 to 2007 the number of Communications submitted to *AC-IE* rose from 449 to 5,489[5,7] (and this trend is continuing); thus when reviewing Communications the editors are under considerable pressure to reach rapid publication decisions.

the editors are under considerable pressure to reach rapid publication decisions

Review of Communications at AC-IE

A Communication submitted to *AC-IE* is usually subject to internal and external review. First, editors at the journal evaluate whether the Communication contributes to the development of an important area of research (internal review).[11] Currently, nearly 20% of submitted Communications are rejected by the editors directly at this stage. If the editors find that a Communication is an appropriate contribution, it goes out to independent referees (external review),[6] who review it and send in a report using an evaluation form and a comment sheet.[7] The referees know the authors' identities, but reviews are not signed (single blinding). The referees are asked to return their reviews to the editorial office within two weeks. This is a very short time period compared to the deadlines given to referees by other journals; a survey of editors of various specialist journals found that the deadline for referees is usually 3–4 weeks.[12] *AC-IE* therefore tries to guarantee rapid reviewing of submitted Communications.

In 2000 the *AC-IE* evaluation form for referees contained the following four questions (in 2008 this was changed to five questions):

1. 'How important do you consider the results?' (four response categories: very important, important, less important, unimportant)
2. 'Do the data obtained by experiment or calculation verify the hypothesis and conclusions?' (two response categories: yes, no)
3. 'Is the length of the manuscript appropriate to its contents?' (three response categories: yes, no – the manuscript is too short, no – the manuscript is too long)
4. 'Do you recommend acceptance of the Communication?' (four response categories: yes – without alterations, yes – after minor alterations, yes – but only after major alterations, no)

If referees find a Communication unsuitable for *AC-IE*, they are asked to name another journal in which the study findings might more suitably be published. Once the editors have received the referees' reports, they make the decision to accept or reject a manuscript for publication.

Although the editors give referees extensive guidelines as to the criteria according to

which submitted Communications are to be examined,[13] there is no written policy on how the editors are supposed to make decisions to accept or reject a manuscript in the light of the referees' reports. However, an indication of the editors' decision practices can be found in their rejection letters to authors. These letters regularly contain phrases such as: 'We have been receiving so many manuscripts that we have almost only been accepting those with two clear-cut referee reports', or 'We have to reject almost all manuscripts that do not receive two clear-cut recommendations.' However, the letters do not provide a clear explanation of what the editors mean by 'clear-cut'; there are only hints, such as the indication found in the following excerpt:

> Regrettably we can currently only accept those papers that have two clear-cut positive recommendations from referees, and in rare cases we even have to decide against these authors (our rejection rate has risen to about 60%). Even though referee [X] deems your results important and certainly worth publishing, he does not recommend publication in *Angewandte Chemie*. Therefore, your paper was not rejected solely on the grounds of referee [Y] [who rated the results as unimportant and recommended rejection of the manuscript].

In the comment on referee X, the editor is referring to the two key questions on the evaluation form: 'How important do you consider the results?' and 'Do you recommend acceptance of the Communication?'

From these and similar indications in the editors' letters, it seems that in general a Communication is published in *AC-IE* only if two referees rate the results of the study as important or very important and also recommend publication in the journal. Thus, the clear-cut rule seems to be that a Communication is published only if two referees choose the response category 'very important' or 'important' to the question, 'How important do you consider the results?' and also do not answer 'no' to the question, 'Do you recommend acceptance of the Communication?' The editors appear to deviate from this rule in only a few cases (see quotation above). If the referees have strongly conflicting opinions as to the importance of a submission (such as 'very important' and 'less important'), the editors ask a top adviser to review both the manuscript and the initial referees' reports.

Database for the present study

For the investigation of manuscript review at *AC-IE* we were given access to information on all 1,899 Communications that were reviewed in the year 2000. The information was taken from archived material that was stored electronically by the publisher, Wiley-VCH. Of the 1,899 Communications reviewed, 878 (46%) were accepted for publication, and 1,021 (54%) were rejected. For 375 manuscripts a late referee report arrived at the editorial office after the editorial decision had been made. In 270 (72%) of these cases two external reviews had already been received; in the remaining 105 cases the sequence of events was more complicated (initial reviews, review of a revised manuscript, review by a top adviser, and so on).

for 375 manuscripts a late referee report arrived after the editorial decision had been made

We carried out statistical analysis of the question of what would have happened if the editorial decision had been based on a late referee report in place of either of the earlier ones, by looking at all Communications for which there was a complete set of answers by all three referees to both of the following two questions on the evaluation form: 'How important do you consider the results?' and 'Do you recommend acceptance of the Communication?' With only a few exceptions, referees that reviewed a revised manuscript or an author's appeal against rejection, and referees that were called in as top advisers, did not answer these two questions on the evaluation form; they used the comment sheet instead. In addition, many of the referees that provided the initial reviews did not answer the two questions. For this reason, only 162 of the 375 Communications mentioned above could be included in the statistical analysis of this study. For these 162 submissions, there were two referee reports, in the light of which an editor made a decision, plus a third report that arrived at the editorial office only after the publication decision had already been made. For the

other 213 Communications with a late report, the prerequisites for the statistical analysis were not met.

Length of time between editorial decision and the arrival of the late report at the editorial office

On average, the 162 third reports arrived at the editorial office approximately one week after a final decision had been made (mean = 1.16 weeks, SD = 1.86, min = 0, max = 15). In many cases, a third report on a Communication was forwarded to the author of the manuscript – regardless of whether it supported the final decision or not. This is the usual practice if the report contains good suggestions for improving the manuscript (before it is submitted by the authors for publication to another journal).

it would be desirable to investigate the same question for other journals

Statistical measure for agreement between referees

To assess the level of agreement between two referees' ratings of the same Communication, Cohen's kappa (κ_c) was used. κ_c is a statistical measure of level of agreement between two or more raters.[14] If the raters are in complete agreement, then $\kappa_c = 1$; if κ_c is near 0, the observed level of agreement is not much higher than a chance level. According to the guidelines of Fleiss,[15] $\kappa_c < 0.4$ indicates poor agreement and $\kappa_c > 0.75$ excellent agreement; a figure between the two indicates good agreement: 'Multiplied by 100, κ indicates the percentage by which two raters' agreement exceeds the agreement that could be expected from chance' (p. 5).[14] If the two-sided *P*-value for κ_c is <0.05 (i.e. if κ_c is statistically significantly different from zero), the null hypothesis of purely random variation is rejected. The coefficient κ_c has the problem 'that the maximum degree of agreement is reduced when raters use the categories at different rates'.[14] Brennan and Prediger[16] propose a variant of κ, named κ_n, to solve this problem. In this study, both variants of κ are used to measure referees' agreement.

Limitations

This study did have some limitations. It covers only 162 submitted manuscripts, which represents only a small proportion (8%) of the 1,899 submitted Communications in the comprehensive dataset. To test the generalizability of the findings, it would be desirable to investigate the same question for other journals. Another limitation of this study might be that it downplays the undoubted role of the *AC-IE* editor's own judgement in difficult cases.

Results

The following section presents the results on agreement among the three referees; in the subsequent section we turn to the 'clear-cut' rule when the editorial decision is based on the first and second referee report, and thirdly we seek an answer to the question of what would have happened if the editorial decision had been based on the third report in place of either the first or the second.

Agreement between the first, second, and third referee

All of the referees' reports (numbered in

Table 1. Agreement between the responses of the first, second, and third referee when the third report arrived at the editorial office after the decision was made

Referee pair	No. of Communications reviewed in common	Observed agreement (%)	Expected agreement (%)	Brennan and Prediger's kappa (κ_n)	Cohen's kappa (κ_c)	95% confidence interval for κ_c[a]
First, second referee	162	78	50	0.56	0.56*	0.43–0.68
First, third referee (i.e. late report)	162	61	50	0.21	0.21*	0.06–0.36
Second, third referee (i.e. late report)	162	62	50	0.24	0.24*	0.09–0.38

[a]Based on 1,000 bootstrap replications for estimates of standard errors.
*$P < 0.05$ (only calculated for κ_c).

Table 2. Agreement between the responses of the first, second, and third referee when all three reports arrived on time

Referee pair	No. of Communications reviewed in common	Observed agreement (%)	Expected agreement (%)	Brennan and Prediger's kappa (κ_n)	Cohen's kappa (κ_c)	95% confidence interval for κ_c[a]
First, second referee	232	52	51	0.04	0.02	–0.11–0.15
First, third referee	232	63	50	0.27	0.27*	0.14–0.39
Second, third referee	232	59	50	0.18	0.18*	0.06–0.31

[a]Based on 1,000 bootstrap replications for estimates of standard errors.
*$P < 0.05$ (only calculated for κ_c).

order of their arrival at the editorial office) gave a positive (+) rating to about half of the 162 Communications (Referee report 1 (R1): 51%, R2: 53%, and R3: 49%). Their overall acceptance rates were thus about the same – and, interestingly, about 50%. Table 1 shows the κ coefficients describing agreement between the ratings of two referees per Communication. As the table shows, the κ coefficients range from 0.21 to 0.56. In other words, the referees show agreement in their ratings of 21–56% more manuscripts than would have been expected by chance. According to guidelines for the interpretation of κ,[15] the coefficients in Table 1 indicate a low level of agreement between R1 and R3 ($\kappa_n = 0.21$) and between R2 and R3 ($\kappa_n = 0.24$), but a good level of agreement between R1 and R2 ($\kappa_n = 0.56$).

This is an interesting result. Why is it that R3 (the late report) tends to disagree with the earlier reviews, while the earlier reviews tend to agree with each other? To find possible reasons for this difference, we tested agreement among R1, R2, and R3 for those Communications in the database of this study for which the editor had at his disposal three reports on which to base the publication decision (385 of the total 1,899 Communications). For 232 of these, all three referees answered both of the two key questions on the evaluation form (concerning importance of results and publication recommendation). Here too, all of the referees' reports gave a positive (+) rating to about half of the 232 Communications (R1: 56%, R2: 57%, and R3: 50%). However, the κ coefficients show a low level of agreement (according to the guidelines of Fleiss,[15,16] between all three pairs of referees' ratings: R1 and R2 $\kappa_n = 0.04$; R1 and R3 $\kappa_n = 0.27$; R2 and R3 $\kappa_n = 0.18$ (see Table 2). In particular, the agreement between R1 and R2 is extremely low.

How can this finding be interpreted? It suggests that the editor waits for a third report if the first two referees' responses to the two key questions do not agree. However, if they do agree, then the editor makes the decision based on the two reports available and does not wait for the third.

the editor waits for a third report if the first two referees' responses do not agree

The editor's application of the 'clear-cut' rule with two reviews

Table 3 shows the relationship between the ratings of the two referees that were taken into consideration for the editor's decision to accept or reject the manuscript for publication, and the referee's rating in the third report. As Table 3 shows, the editors of *AC-IE* did indeed publish only those manuscripts that – following the 'clear-cut' rule described above – were rated positively (+) by both R1 and R2: 65 manuscripts were rated positively (+) by both referees and accepted for publication; 93 manuscripts were rated negatively (–) by one or both referees and rejected by the editors.

There were only four cases where the editor did not adhere to the 'clear-cut' rule. These few deviations can be readily explained, however, based on further information contained in the archive material. One of the manuscripts was rejected for publication even though both referees had previously rated it positively (+); the manuscript was forwarded to *AC-IE*'s sister journal *ChemPhysChem* (*A European Journal of Chemical Physics and Physical Chemistry)*, in which it was published. In this case, the

Table 3. Referee's rating in a third (late) report where the editor has already made a publication decision based on the first and the second referees' ratings

First referee's rating	Second referee's rating	Editor's final decision	Third referee's rating		
			Total	+	–
+	+	Acceptance	65	42	**23**
–	+	Acceptance	3	1	2
+	+	Rejection	1	1	0
+	–	Rejection	16	**6**	10
–	+	Rejection	17	**8**	9
–	–	Rejection	60	22	38
		Total	162	80	82

The symbol '+' in the table means that a referee answered 'very important' or 'important' to the question, 'How important do you consider the results?', and did not answer 'no' to the question, 'Do you recommend acceptance of the Communication?'

The symbol '–' indicates that a referee answered 'less important' or 'unimportant' to the question, 'How important do you consider the results?', and/or answered 'no' to the question, 'Do you recommend acceptance of the Communication?'

The four cases in the table framed with a dotted line are cases where the editor's decision did not follow the 'clear-cut' rule, as explained in the text.

The bold figures indicate those Communications where the third report would have changed the editorial decision under the 'clear-cut' rule.

the editors followed the 'clear-cut' rule in all but a few cases, in reaching a publication decision

editor agreed with a referee's opinion that the results of the study reported in the Communication were important and definitely deserved publication but that the topic was too specialized in nature considering the general scope of *Angewandte Chemie*. In the editor's opinion, the manuscript would make an ideal contribution to *ChemPhysChem*. In addition, three manuscripts did not receive clear-cut recommendations from both R1 and R2 but were nonetheless published in *AC-IE*. Closer inspection of the referee reports reveals that in these cases, one referee gave the manuscript an extremely positive rating, and the second referee's criticism consisted in the recommendation that the manuscript was more suitable for publication in a specialized journal or as a full paper rather than in the form of a Communication. In the end, the editors did not follow the second referee's recommendations.

Altogether, the results in Table 3 indicate that the editors followed the 'clear-cut' rule in all but a few cases, in reaching a publication decision. In order to validate this result, we examined the editors' decision practices on Communications for which the editor based his decision on two referees' reports and where there was *no* third report after the decision had been made. Did the editors also follow the clear-cut rule for these submissions? For 949 (50%) of the 1,899 manuscripts submitted there were two referees' reports available, in the light of which the editor made the decision. For 529 of these, the two referees answered both of the two key questions outlined above. In only 37 cases (7%) did an editor *not* follow the 'clear-cut' rule. We can therefore assume that the 'clear-cut' rule is the editors' common practice when making the publication decision based on two referee reports.

What would have happened if the editorial decision had been based on the third referee report in place of either the first or the second?

To answer this question we analyzed 158 cases where the editors followed the 'clear-cut' rule when deciding to accept or reject a manuscript for publication (see Table 3). The analysis described below is therefore based on 98% of the cases – i.e. not including the four cases where the editor did not follow the 'clear-cut' rule. For how many of these Communications – according to the 'clear-cut' rule – would a different decision have been expected? As the results in Table 3 show, the editors might have made a

Table 4. Referees' ratings where the editor based the decision whether or not to publish a Communication on three reviews

First referee's rating	Second referee's rating	Third referee's rating	Editor's final decision		
			Acceptance	Rejection	Total
+	+	+	51	**4**	55
–	+	+	20	**2**	22
+	+	–	15	**5**	20
+	–	+	21	**4**	25
+	–	–	**1**	28	29
–	–	+	**2**	12	14
–	+	–	**3**	33	36
–	–	–	0	31	31
			113	119	232

The symbol '+' in the table means that a referee answered 'very important' or 'important' to the question, 'How important do you consider the results?', and did not answer 'no' to the question, 'Do you recommend acceptance of the Communication?'

The symbol '–' indicates that a referee answered 'less important' or 'unimportant' to the question, 'How important do you consider the results?', and/or answered 'no' to the question, 'Do you recommend acceptance of the Communication?'

The bold figures in the table are cases where the editor's decision did not follow the general rule, as explained in the text. In the four cases where the manuscript was rejected with three positive (+) ratings, the authors retracted the manuscript.

different decision on 37 (23%) of the manuscripts. On 23 manuscripts (15%), instead of two clear-cut recommendations (+ +), there would have been one positive (+) and one negative (–) rating by the referees; rather than to accept, the editors would most likely have rejected the manuscript. On 33 manuscripts (21%), instead of a positive (+) and negative (–) rating, there would have been two (positive) 'clear-cut' recommendations in 14 cases (1%) if the positive (+) one of the original pair was chosen; these manuscripts would have been accepted rather than rejected; on the other hand, if the negative (–) one was chosen, the late review would still have resulted in rejection. Similarly, the negative ratings (–) received by the remaining 19 manuscripts in this group would not have affected the publication decision. For the 60 manuscripts rejected because the ratings in both R1 and R2 were negative (–), the rating in the third report (R3) was positive (+) in 22 cases, but this was unlikely to have changed the editorial decision; neither would it have changed in the remaining 38 cases when the third report was also negative (–). In other words, the editor's decision on 37 (2%) of the 1,899 submitted manuscripts might have been different if late reports had been received earlier.

As mentioned above, for 385 of the total 1,899 submissions, the editors based the decision whether or not to publish not on two but on three reviews. Having three referee reports prior to editorial decision is the second most frequent situation in the *AC-IE* database; the most frequent is two reports. For the 158 cases on which an editor has made a decision based on the 'clear-cut' rule (see Table 3) we also tested what would have happened if the third (late) referee report had been considered in the editorial decision in addition to (and not instead of) the first and second referee reports. This question can be answered only if the editors also followed a general rule when three referee reports were available. As the results for 232 of the 385 manuscripts show, when all three referees answered both of the key questions mentioned above, generally those submissions that received a positive (+) rating from at least two referees were accepted for publication (see Table 4). Only in the case of 21 Communications (9%) did the editors deviate from this rule (in many cases because another journal was seen as more suitable for publication of the manuscript). This means that, when three reviews are available, one negative (–) and two positive (+) reviews generally result in acceptance for publication, while one positive (+) and two negative (–) reviews lead to rejection (clearly, with three negative (–) reviews, the

generally those submissions that received a positive rating from at least two referees were accepted for publication

decision is not in question – this was the case in 31 instances).

Applying this rule to the 158 manuscripts in the database of this study (see Table 3), for the 33 that were rejected for publication based on one positive (+) and one negative (–) rating, a different decision would have been expected in 14 cases, since the third rating was positive (+); the remaining 19 would have been unaffected, since the rating was negative (–). For the 65 cases accepted by the editor with two positive (+) ratings (see Table 3), the third report would not have changed the decision. Whether the third report was positive (+) or negative (–), they, too, would be accepted for publication. Similarly, the recommendation of the third report could not affect the outcome in the 60 cases where the decision was made to reject based on two negative (–) ratings.

random elements, such as the order in which referees submit their reports, should not play a role in peer-review decisions

Discussion

The present study examined the question of what would have happened if the editorial decision had been based on a third referee report in place of either the first or the second referee report. As the editors at *AC-IE* follow a 'clear-cut' rule in their decisions on most submissions (i.e. a submission is accepted for publication only if the ratings of two referees are positive (+), as defined above), using the data from an extensive research project on the peer-review process at *AC-IE* we were able to seek an answer to the question as to the luck of the referee draw. As the results of this study show, the 'clear-cut' rule could have led to a different editorial decision for 23% of the submissions for which there was a third (late) referee report (37 of 162) – i.e. for about 37 (2%) of the total 1,899 submitted manuscripts.

Random elements, such as the order in which referees submit their reports, should not play a role in peer-review decisions. The editors at *AC-IE* as a rule send out a submitted manuscript to three referees, even though for the 'clear-cut' rule in most cases only two referee reports are used as a basis for the editorial decision. While requesting three referees to review a manuscript but not using all of the reports for the editorial decision may speed up the editorial process (and speed is very important when it comes to publishing Communications), the editors run the risk that a third, late, referee report would have suggested a different decision. However, if a manuscript is rejected for publication and the editors receive a late positive report, *AC-IE* does grant authors the option to appeal the decision.

With the very many manuscripts that are submitted to *AC-IE*, using a peer-review system that counts referee 'votes' (as in the 'clear-cut' rule here) is probably what makes the review process at all manageable. However, this practice could be seen as not following the best peer-review practices. Ideally, editors should not simply go along with referees' recommendations on whether to publish or not.[2] Low inter-referee agreement about a given manuscript is a common problem of the journal peer-review system.[17] According to Fletcher and Fletcher,[18]

> the opinions of two reviewers, even if chosen at random from all possible reviewers, are too few in themselves to yield a statistically stable basis for deciding whether or not the manuscript should be published. Indeed, one would need to have at least six reviewers, all favouring publication or rejection, for their votes to yield a statistically significant conclusion ($P < 0.05$). (p. 66)

Editors may also select reviewers because they expect divergent opinions – in which case 'counting votes' makes no sense: 'If reviewers are advisers to editors, then that advice is richer if their reviews reflect different expertise and values, and as a result disagree on the overall strength of the manuscript.'[18]

As the data on peer review at *AC-IE* show, the editors follow the 'clear-cut' rule in most but not all cases. With some submissions, the editors wait for a third referee report, even if two referee reports have already come in, or a top adviser may be consulted in a very controversial case. However, no written policy exists as to what review route is to be followed in what cases. For instance, if referee reports do not agree, when is the editorial decision made according to the 'clear-cut' rule, and when is a top adviser called in? For transparency and fair-

ness of the peer-review process at *AC-IE*, it would be very useful to have a written policy on the possible review steps and the rules applied. If all referees knew of the existence of the 'clear-cut' rule, for example, they would surely be conscious of their responsibility to send in their reports to the editors within the journal's two-week deadline.

Acknowledgements

The entire research study, which is also investigating quality assurance of open access journals, is supported by a grant from the Max Planck Society. We thank Dr. Peter Gölitz, Editor-in-Chief of *Angewandte Chemie*, the Editorial Board of *Angewandte Chemie*, and the German Chemical Society (GDCh, Frankfurt am Main, Germany) 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 carrying out of the study. We thank Dr Rex Galbraith (University College London) and Dr Rüdiger Mutz (ETH Zurich) for their help in undertaking the statistical analyses.

References

1. Braun, T. Keeping the gates of science journals. Gatekeeping indicators of national performance in the sciences. In Moed H.F., Glänzel W. and Schmoch U., editors. *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*. Dordrecht, Kluwer, 2004, pp. 95–114.
2. Hames, I. *Peer Review and Manuscript Management of Scientific Journals: Guidelines for Good Practice*. Oxford, Blackwell, 2007.
3. Bornmann, L. and Daniel, H.-D. 2008. Selecting manuscripts for a high impact journal through peer review: a citation analysis of Communications that were accepted by *Angewandte Chemie International Edition*, or rejected but published elsewhere. *Journal of the American Society for Information Science and Technology*, 59: 1841–52. http://dx.doi.org/10.1002/asi.20901
4. Bornmann, L and Daniel, H.-D. 2008. The effectiveness of the peer review process: inter-referee agreement and predictive validity of manuscript refereeing at *Angewandte Chemie*. *Angewandte Chemie International Edition*, 47: 7173–8. http://dx.doi.org/10.1002/anie.200800513
5. Gölitz, P. 2008. Appeals. *Angewandte Chemie International Edition*, 47: 7144–5. http://dx.doi.org/10.1002/anie.200803308
6. Kemsley, J. 2008. Is peer review honest? *Chemical and Engineering News*, 86: 48–9. http://pubs.acs.org/cen/science/86/8606sci1.html
7. Daniel, H.-D. *Guardians of Science: Fairness and Reliability of Peer Review.* Weinheim, Wiley-VCH, 1993. http://dx.doi.org/10.1002/3527602208
8. Merton, R.K. 1957. Priorities in scientific discovery – a chapter in the sociology of science. *American Sociological Review*, 22: 635–59. http://dx.doi.org/10.2307/2089193
9. *Journal of the American Chemical Society*, Guidelines for Authors. Available at: http://pubs.acs.org/paragonplus/submission/jacsat/jacsat_authguide.pdf (accessed 10 July 2008)
10. Neuhaus, C., Marx, W. and Daniel, H.-D. 2009. The publication and citation impact profiles of *Angewandte Chemie* and the *Journal of the American Chemical Society* based on the sections of *Chemical Abstracts*: a case study on the limitations of the Journal Impact Factor. *Journal of the American Society for Information Science and Technology*, 60: 176–83. http://dx.doi.org/10.1002/asi.20960
11. Gölitz, P. 2005. Who is going to read all this? *Angewandte Chemie International Edition*, 44: 5538–41. http://dx.doi.org/10.1002/anie.200502730
12. Association of Learned and Professional Society Publishers. *Current Practice in Peer Review.* Worthing, Association of Learned and Professional Society Publishers, 2000. Available at http://www.alpsp.org/ForceDownload.asp?id = 140
13. *Angewandte Chemie-International Edition*, Guidelines for referees for Communications. Available at http://www3.interscience.wiley.com/journal/26737/home//2002_web_ref.html (accessed 11 June 2008)
14. von Eye, A. and Mun, E. Y. *Analyzing Rater Agreement: Manifest Variable Methods.* Mahwah, NJ, Lawrence Erlbaum Associates, 2005.
15. Fleiss, J. *Statistical Methods for Rates and Proportions.* 3rd edition. New York, Wiley VCH, 2003.
16. Brennan R.L. and Prediger, D.J. 1981. Coefficient kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41: 687–99. http://dx.doi.org/10.1177/001316448104100307.
17. Weller, A.C. *Editorial Peer Review: Its Strengths and Weaknesses.* Medford, NJ, Information Today, Inc., 2002.
18. Fletcher, R.H. and Fletcher, S.W. The effectiveness of journal peer review. In Godlee, F. and Jefferson, T., editors. *Peer review in health sciences*, 2nd edition. London, BMJ Books, 2003, pp. 62–75.

Dr Lutz Bornmann (corresponding author)
Professorship for Social Psychology and Research on Higher Education
ETH Zurich, Zähringerstr. 24
CH-8092 Zurich, Switzerland
Email: bornmann@gess.ethz.ch

Prof. Dr. Hans-Dieter Daniel
Professorship for Social Psychology and Research on Higher Education
ETH Zurich and University of Zurich, Evaluation Office
Mühlegasse 21, CH-8001 Zurich, Switzerland
Email: daniel@evaluation.uzh.ch

it would be very useful to have a written policy on the possible review steps and the rules applied