

## The validity of staff editors' initial evaluations of manuscripts: a case study of *Angewandte Chemie International Edition*

Lutz Bornmann · Hans-Dieter Daniel

Received: 4 January 2010 / Published online: 31 March 2010  
© Akadémiai Kiadó, Budapest, Hungary 2010

**Abstract** This paper investigates the extent to which staff editors' evaluations of submitted manuscripts—that is, internal evaluations carried out before external peer reviewing—are valid. To answer this question we utilized data on the manuscript reviewing process at the journal *Angewandte Chemie International Edition*. The results of this study indicate that the initial internal evaluations are valid. Further, it appears that external review is indispensable for the decision on the publication worthiness of manuscripts: (1) For the majority of submitted manuscripts, staff editors are uncertain about publication worthiness; (2) there is a statistically significant proportional difference in “Rejection” between the editors' initial evaluation and the final editorial decision (after peer review); (3) three-quarters of the manuscripts that were rated negatively at the initial internal evaluation but accepted for publication after the peer review had far above-average citation counts.

**Keywords** Peer review · Validity · Staff editor's initial evaluation

### Introduction

In journal peer review, manuscripts are usually reviewed by several peers before the journal staff editor makes a decision to accept or reject for publication. Some journals “rely on reviewer judgements to a great extent (e.g. always accepting submissions if two, or a majority of, reviewers recommend this)...However, in many journals, although editors base their decision on the reviewers' comments, they do not necessarily follow the reviewers' recommendations about acceptance or rejection” (Wager et al. 2006).

---

L. Bornmann (✉) · H.-D. Daniel  
ETH Zurich, Zähringerstr. 24, 8092 Zurich, Switzerland  
e-mail: bornmann@gess.ethz.ch

H.-D. Daniel  
e-mail: daniel@evaluation.uzh.ch

H.-D. Daniel  
University of Zurich, Evaluation Office, Mühlegasse 21, 8001 Zurich, Switzerland

In this study we investigated the extent to which staff editors are also able to validly assess the quality of submissions without the support of reviewers. To answer this question, we first examined a comprehensive dataset on the manuscript reviewing process at the journal *Angewandte Chemie International Edition* (AC-IE) to see whether the staff editors made valid initial evaluations of a manuscript—after it had been submitted but before external peer review. To examine validity we used the citation counts for the manuscripts later published. In a second step, we examined the number of manuscripts for which there was a change in the editors' evaluations between the initial internal evaluation and their final decision (after peer review).

Citation rates have been a controversial measure of quality (Bornmann and Daniel 2008d). Nevertheless, Lokker et al. (2008) succeeded in demonstrating for clinical articles that publications regarded as important by experts in the appropriate research field shortly after their appearance were cited much more frequently in subsequent years than publications that were less highly regarded. Furthermore, the results of a comprehensive citation content analysis conducted by Bornmann and Daniel (2008b) show that “an article with high citation counts had greater relevance for the citing author than an article with low citation counts” (p. 35).

## Methods

### Manuscript reviewing at AC-IE

AC-IE is one of the prime chemistry journals in the world, with a higher annual Journal Impact Factor (JIF, provided by Thomson Reuters, Philadelphia, PA) than the JIFs of comparable journals (at 10.879 in the 2008 Journal Citation Reports, Science Edition). AC-IE is a journal of the German Chemical Society (Gesellschaft Deutscher Chemiker (GDCh), Frankfurt am Main, Germany) and is published by Wiley-VCH (Weinheim, Germany).

A manuscript submitted to AC-IE usually undergoes internal and external review. First, staff editors at the journal evaluate the quality of the manuscript and determine whether the manuscript contributes to the development of an important area of research (internal review). The editors are fulltime members of the editorial staff. According to Gölitz, the Editor-in-Chief of AC-IE, “the fate of manuscripts and the selection of referees is determined by me, the two Deputy Editors, Neville Compton (Inorganic Chemistry) and Haymo Ross (Organic Chemistry), and four Senior Associate Editors, Frank Maass (Inorganic Chemistry), José Oliveira (Organic Chemistry), Guy Richardson (Inorganic Chemistry), and Diane Smith (Supramolecular Chemistry). All these editors have Doctorates in their specialist areas and in most cases postdoctoral research experience as well as many years experience as full-time editors” (Gölitz 2005, p. 5540). The editors rate the manuscript initially using the following 5-point scale from very positive to very negative: +, ++/–, +/-, +/--, and –. After this initial evaluation by the editor, the submitted manuscript is sent to several independent reviewers (external review), who review it using an evaluation form and a comment sheet. The journal staff editors then make the decision to accept or reject a manuscript for publication on the basis of these reviews and their own evaluations. In the manuscript reviewing process at AC-IE, the staff editor thus evaluates one and the same manuscript twice—once before and once after peer review.

AC-IE introduced peer review in 1982, primarily in conjunction with one of the document types published in the journal, “Communications,” which are short reports on work

in progress or recently concluded experimental or theoretical investigations. What the staff editors of AC-IE look for most of all is excellence in chemical research. Submissions that reviewers deem to be of high quality are selected for publication: For most submissions a manuscript is published only if two external reviewers rate the results of the study reported in the manuscript as (very) important and also recommend publication in the journal (Bornmann and Daniel 2009, 2010). The AC-IE staff editors are thus guided in their publication decisions to a great extent by the reviewers' recommendations.

#### Database for the present study and conducting of citation analysis

For the investigation of manuscript selection at AC-IE we used information on all 1,899 manuscripts that were reviewed in the year 2000. Of the 1,899 manuscripts, 46% ( $n = 878$ ) were accepted for publication in AC-IE, and 54% ( $n = 1,021$ ) were rejected. A search in the literature databases Science Citation Index (SCI) (Thomson Reuters) and Chemical Abstracts (CA) (Chemical Abstracts Services, CAS, Columbus, OH) revealed that of the 1,021 rejected manuscripts, 959 (94%) were later published in 136 other (different) journals.

For accepted and rejected (but published elsewhere) manuscripts, we determined the number of citations for a fixed time window of 3 years after the publication year. "Fixed citation windows are a standard method in bibliometric analysis, in order to give equal time spans for citation to articles published in different years, or at different times in the same year" (Craig et al. 2007, p. 243). The citation analyses for the present study were conducted in the year 2007 based on CA. CA is a comprehensive database of publicly disclosed research in chemistry and related sciences (see <http://www.cas.org/>).

Of all 1,837 manuscripts published in the AC-IE (accepted manuscripts) or another journal (rejected manuscripts, see above), 1,817 could be included in the citation analysis. For 20 rejected manuscripts that were published elsewhere between 2004 and 2006, a citation window of 3 years (1 year after publication up to the end of 2006) was not available.

The statistical analyses of the citations in Bornmann and Daniel (2008a, c) show that the manuscripts accepted by AC-IE were on average cited statistically significant more frequently than manuscripts rejected by the journal and published elsewhere. This finding provides evidence for the predictive validity of the editorial publication decisions at AC-IE.

#### Statistical analysis

The assumption of a negative binominal distribution of citation counts (Allison 1980) makes it necessary to calculate a negative binomial regression model to test for differences between citation counts for manuscripts that are assessed differently by the AC-IE staff editor (before peer review) (Bornmann et al. 2008; StataCorp 2007). A posteriori contrasts with Bonferroni correction were tested in order to determine what pair-wise differences in mean citations counts are statistically significant.

In order to test whether there was a change in the evaluation category (from accept to reject) for a statistically significant number of the manuscripts between the initial internal evaluation by the AC-IE editor (before peer review) and the editorial decision (after peer review), the McNemar (1947) chi-square test was calculated. The McNemar test is used when two dependent samples (here, manuscripts assessed before and after peer review) are compared regarding a dichotomous variable (here, accept and reject).

**Results**

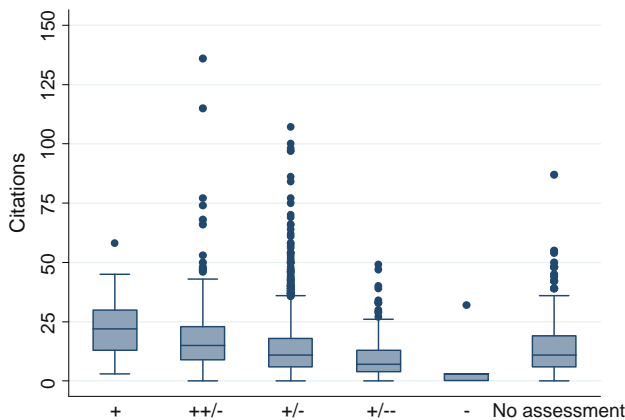
Table 1 shows the distribution of the initial internal evaluations by the AC-IE staff editors. The editors rated 13% of the manuscripts positively, choosing one of the two positive categories +(1%) or ++/- (12%); they assigned 8% of the manuscripts to the negative category +/--. Only 8 manuscripts were assigned to the very negative category (-). Approximately two-thirds of the manuscripts (67%) were not rated either positively or negatively by the staff editor before peer review; these manuscripts were assigned the rating +/--. The high number of manuscripts in +/- category indicates that in many cases, the editors are uncertain as to the publication worthiness of a manuscript and make the decision to accept or reject for publication based on the external reviewers' recommendations (see Methods section).

As Fig. 1 shows, the initial internal evaluations by the AC-IE staff editors correspond with the later citation counts for all submissions accepted by AC-IE or rejected by AC-IE but published elsewhere ( $n = 1,800$ ): Manuscripts that the editors rated the highest before peer review (+,  $n = 25$ ), were cited 22 times on average (median). Manuscripts that the

**Table 1** Initial internal evaluations by the AC-IE staff editors

| Evaluation by AC-IE editors | Absolute numbers | In percent |
|-----------------------------|------------------|------------|
| +                           | 26               | 1          |
| ++/-                        | 227              | 12         |
| +/-                         | 1,256            | 67         |
| +/--                        | 147              | 8          |
| -                           | 8                | 0          |
| No assessment               | 218              | 12         |
| Total                       | 1,882            | 100        |

*Notes.* In the category 'No assessment' are manuscripts for which the staff editor entered no rating on the evaluation sheet



**Fig. 1** Box plots for citations of manuscripts that received different initial internal evaluations by an AC-IE staff editor ( $n = 1,800$ :  $n = 25$  (+),  $n = 222$  (++/-),  $n = 1,201$  (+/-),  $n = 133$  (+/--),  $n = 7$  (-), and  $n = 212$  with no assessment). The horizontal line in the middle of each box indicates the median, and the top and bottom borders of the box mark the 75th and 25th percentiles, respectively. The whiskers above and below the box mark the upper and lower adjacent values. The points above the whiskers are defined as outliers

staff editors rated  $++/-$  ( $n = 222$ ) were cited 15 times on average (median). The other average citation rates (median) for the differently rated manuscripts are:  $+/-$  ( $n = 1,201$ ) = 11,  $+/--$  ( $n = 133$ ) = 7, and  $-$  ( $n = 7$ ) = 3. The result of the negative binomial regression models shows that whereas the mean citation rates for the manuscripts rated  $+$  and  $++/-$  do not differ statistically significantly ( $P = .22$ ), the differences between the mean citation rates of all of the other manuscript groups are statistically significant ( $P < .05$ ). These findings indicate that the initial evaluations of the manuscripts by the editors can be called valid. The analyses concerning the validity of the editors' final decisions (after peer review) showed similar results: accepted manuscripts were cited more frequently on average (median = 14) than rejected (but published elsewhere) manuscripts (median = 9) (Bornmann and Daniel 2008a, c; Bornmann et al., submitted).

Table 2 shows the connection between initial internal evaluation and editor's final decision. Of the manuscripts that the editor evaluated positively ( $+$  or  $++/-$ ,  $n = 253$ ) before peer review, 199 (79%) were later accepted for publication, and 54 (21%) were rejected. Of the manuscripts that the staff editor evaluated negatively ( $+/--$  or  $-$ ,  $n = 155$ ) before peer review, 134 (86%) were later rejected, and 21 (14%) were accepted. Thus, for the majority of manuscripts that the editor rated positively or negatively upon the initial internal evaluation, the initial internal evaluations and final decisions correspond. (Only those manuscripts were included in the analysis that the staff editor had evaluated positively or negatively before peer review—that is, manuscripts rated in the category  $+/-$  ( $n = 1,256$ ) were *not* included.)

To check whether a significant number of manuscripts were assigned a different rating between the initial internal evaluation and the final decision, the McNemar chi-square test was computed. The data in Table 3 are based on the McNemar test. So that uniform categories at both time points would be available for the analysis, the evaluation categories  $+$  and  $++/-$  were combined in a category called 'accept,' and the evaluation categories  $+/--$  and  $-$  were combined in a category called 'reject'. As the result of the McNemar chi-square test in Table 3 shows, that the marginal proportions are statistically significant from each other. Thus, there is a statistically significant proportional difference in "Rejection" between the initial internal evaluation and the final decision.

**Table 2** AC-IE staff editor's final decision (after peer review) by the editor's initial internal evaluation (before peer review)

| Initial internal evaluation | Final decision |           | Total |
|-----------------------------|----------------|-----------|-------|
|                             | Acceptance     | Rejection |       |
| $+$                         | 23             | <b>3</b>  | 26    |
| $++/-$                      | 176            | <b>51</b> | 227   |
| $+/-$                       | 554            | 702       | 1,256 |
| $+/--$                      | <b>21</b>      | 126       | 147   |
| $-$                         | 0              | 8         | 8     |
| No assessment               | 101            | 117       | 218   |
| Total                       | 875            | 1,007     | 1,882 |

*Notes.* The numbers shown in bold typeface are manuscripts for which the staff editors' final decision after peer review differed from their initial evaluation before peer review. The row in the table shown with italics are those manuscripts for which the editors made no positive or negative rating before peer review—that is, those manuscripts about which the editors were undecided

**Table 3** AC-IE staff editor's final decision (after peer review) by the editor's initial internal evaluation (before peer review)

| Initial internal evaluation | Final decision |           | Total |
|-----------------------------|----------------|-----------|-------|
|                             | Acceptance     | Rejection |       |
| Acceptance (+ or ++/–)      | 199            | 54        | 253   |
| Rejection (+/– or –)        | 21             | 134       | 155   |
| Total                       | 220            | 188       | 408   |

Note.  $\chi^2_3 = 14.5$ ,  $P < .05$

As a final step, for the manuscripts that were assigned a different evaluation by the staff editor after peer review, we examined which of the two evaluations (initial evaluation of the manuscript before peer review and final decision after peer review) of the individual submission was the more valid evaluation, using ex-post determined citation counts for the manuscripts as a quality indicator. Following van Raan (2004), a positive evaluation by the AC-IE editor was called valid if the manuscript was cited a far above-average number of times after publication (and vice versa). (As a yardstick we did not choose above average but instead *far* above average, because the AC-IE seeks to publish excellence in chemical research.) For this we compared the citation count of an individual manuscript with the average citation rate for papers published in the same chemical sub-field (CA section) (Bornmann and Daniel 2008c). If the citation count was higher than 1.5 times of the mean value, we called the impact of the manuscript far above average.

Of the 54 manuscripts that the staff editor evaluated positively for the initial internal evaluation but rejected in the final decision (see Table 3), 48% had far above-average citation counts (eight manuscripts could not be included in the analysis, as no citations were available; see Methods section). Of the 21 manuscripts (see Table 3) that the editor initially evaluated negatively but were later accepted for publication, 75% had far above-average citation counts (in this analysis, one manuscript could not be included due to missing citations). Thus, for the majority of these manuscripts, the staff editors' assessments are more valid *after* peer review (that is, in the final decision).

## Discussion

The results of this study indicate that the initial internal evaluations are valid. Further, it appears that external review is indispensable for the decision on the publication worthiness of manuscripts: (1) For the majority of submitted manuscripts, staff editors are uncertain about publication worthiness; (2) there is a statistically significant proportional difference in "Rejection" between the editors' initial evaluation and the final editorial decision (after peer review); (3) three-quarters of the manuscripts that were rated negatively at the initial internal evaluation but accepted for publication after the peer review had far above-average citation counts.

All in all, peer review increases the validity of the publication decisions by the journal staff editors.

**Acknowledgements** The broader research study, which is also investigating quality assurance of open access journals, is supported by a grant from the Max Planck Society. The authors would like to thank Dr. Christophe Weymuth (formerly at the Organic Chemistry Institute of the University of Zurich and now at

BIOSYNTH AG, Switzerland) for investigation of the manuscripts rejected by *Angewandte Chemie International Edition* and published elsewhere. We also thank Dr. Werner Marx and Dr. Hermann Schier of the Central Information Service for the institutes of the Chemical Physical Technical (CPT) Section of the Max Planck Society (located at the Max Planck Institute for Solid State Research in Stuttgart, Germany) for conducting the search for citations of the accepted and rejected (but published elsewhere) manuscripts in the literature database Chemical Abstracts (CA). We thank Dr. Peter Göllitz, 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 selection process of the journal and thank the members of the editorial office for their generous support during the carrying out of the study.

## References

- Allison, P. D. (1980). Inequality and scientific productivity. *Social Studies of Science*, *10*(2), 163–179.
- Bornmann, L., & Daniel, H.-D. (2008a). The effectiveness of the peer review process: Inter-referee agreement and predictive validity of manuscript refereeing at *Angewandte Chemie. Angewandte Chemie International Edition*, *47*(38), 7173–7178.
- Bornmann, L., & Daniel, H.-D. (2008b). Functional use of frequently and infrequently cited articles in citing publications. A content analysis of citations to articles with low and high citation counts. *European Science Editing*, *34*(2), 35–38.
- Bornmann, L., & Daniel, H.-D. (2008c). Selecting manuscripts for a high impact journal through peer review: a citation analysis of Communications that were accepted by *Angewandte Chemie International Edition*, or rejected but published elsewhere. *Journal of the American Society for Information Science and Technology*, *59*(11), 1841–1852.
- Bornmann, L., & Daniel, H.-D. (2008d). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80.
- Bornmann, L., & Daniel, H.-D. (2009). The luck of the referee draw: The effect of exchanging reviews. *Learned Publishing*, *22*(2), 117–125.
- Bornmann, L., & Daniel, H.-D. (2010). The manuscript reviewing process—empirical research on review requests, review sequences and decision rules in peer review. *Library & Information Science Research*, *32*(1), 5–12.
- Bornmann, L., Mutz, R., Marx, W., Schier, H., Daniel, H.-D. (submitted). A multilevel modelling approach to investigating the predictive validity of editorial decisions: Do the editors of a high-impact journal select manuscripts that are highly cited after publication?
- Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H.-D. (2008). Use of citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, *8*, 93–102.
- Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact? A critical review of the literature. *Journal of Informetrics*, *1*(3), 239–248.
- Göllitz, P. (2005). Who is going to read all this? *Angewandte Chemie International Edition*, *44*(35), 5538–5541.
- Lokker, C., McKibbin, K. A., McKinlay, R. J., Wilczynski, N. L., & Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: retrospective cohort study. *British Medical Journal*, *336*(7645), 655–657.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*, 153–157.
- StataCorp. (2007). *Stata statistical software: release 10*. College Station, TX, USA: Stata Corporation.
- van Raan, A. F. J. (2004). Measuring science. Capita selecta of current main issues. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems* (pp. 19–50). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Wager, E., Parkin, E., & Tamber, P. (2006). Are reviewers suggested by authors as good as those chosen by editors? Results of a rater-blinded, retrospective study. *BMC Medicine*, *4*(1), 13.