Scientometrics
DOI 10.1007/s11192-012-0731-8

In public peer review of submitted manuscripts, how do reviewer comments differ from comments written by interested members of the scientific community? A content analysis of comments written for *Atmospheric Chemistry and Physics*

Lutz Bornmann · Hanna Herich · Hanna Joos · Hans-Dieter Daniel

Received: 16 November 2011

Abstract Whereas in traditional peer review a few selected researchers (peers) are included in the manuscript review process, public peer review includes both invited reviewers (who write 'reviewer comments') and interested members of the scientific community who write comments ('short comments'). Available to us for this investigation are 390 reviewer comments and short comments assessing 119 manuscripts submitted to the journal *Atmospheric Chemistry and Physics* (ACP). We conducted a content analysis of these comments to determine differences in the main thematic areas considered by the scientists in their assessment comments. The results of the analysis show that in contrast to interested members of the scientific community, reviewers focus mainly on (1) the formal qualities of a manuscript, such as writing style, (2) the conclusions drawn in a manuscript, and (3) the future "gain" that could result from publication of a manuscript. All in all, it appears that 'reviewer comments' better than 'short comments' by interested members of the scientific community support the two main functions of peer review: selection and improvement of what is published.

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

H. Herich · H. Joos
Institute of Atmospheric and Climate Science, ETH Zurich,
Universitätstrasse 16, 8092 Zurich, Switzerland

Present Address:
H. Herich
Swiss Federal Laboratories for Materials Science and Technology,
Empa, Überlandstrasse 129, 8600 Dübendorf, Switzerland

H.-D. Daniel
Professorship for Social Psychology and Research on Higher Education,
ETH Zurich, Mühlegasse 21, 8001 Zurich, Switzerland

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

Published online: 25 April 2012

Keywords Public peer review · Content analysis · Thematic areas for manuscript review · Criteria and reasons in manuscript review

Introduction

Already in the late 1990s the American Association for the Advancement of Science (1998) pointed out the importance of peer review and adequate quality control for electronic publication (Bornmann 2011). With public peer review (PPR), electronic publishing offers new possibilities of quality assurance that cannot be realized in traditional closed peer review. Whereas traditional peer review of submitted manuscripts involves the use of designated reviewers (peers selected by editors) who are anonymous to the authors, open peer review involves reviewers who are not anonymous to the authors. PPR—in addition to invited reviewers—is open to a wider circle of scientists who are interested in a manuscript and wish to comment on it. The idea behind PPR is to make possible expert discussions of individual scientific contributions. The discussions are to be held publicly and are to be transparent and accessible afterwards.

The advantage of PPR is considered to be that with the aid of comments by interested members of the scientific community, more valid publications decisions can be made and accepted manuscripts can be better revised/improved prior to publication than is the case with traditional close peer review, which is based exclusively on reviewer comments. Furthermore, it is expected that public peer review increases the chance of preventing the publication of falsified results (see here Xin 2006). Finally, "publishing the views of dissenting referees or the contrasting views of qualified commentators would allow readers an opportunity to judge a manuscript's merits on their own and also provide protection to naive readers" (Bedeian 2004, p. 211). However, besides the advantages of PPR, researchers have pointed out a number of disadvantages. Harnad (2000) raised the question as to whether self-appointed commentators can in fact be expected to make qualified comments: "The expert population in any given speciality is a scarce resource, already overharvested by classical peer review, so one wonders who would have the time or inclination to add journeyman commentary services to this load on their own initiative" (p. 9). McCormack (2009) doubted whether scientists are willing to comment publicly on the work of others: "It may be unrealistic to expect individuals in the same field to critique each other openly and honestly. Those researchers who do advocate for the process acknowledge that junior members in a field would be afraid to openly criticize more senior and established members without hurting their own career opportunities. They might even be pressured to write favorably about work they don't support" (p. 68, see here also Mizzaro 2003).

Initiatives to open up peer review started in 1959, when open peer review was first implemented by the journal *Current Anthropology* (Daniel 1993). The journal *Behavioral and Brain Sciences*, founded in 1978 modelled on *Current Anthropology*'s open peer commentary feature (Harnad 1978). The *Journal of Molecular and Cellular Immunology*, founded in 1983, also practiced an open peer review policy—but it was forced to cease publication in 1990 (Janeway 1990). Further initiatives which have been attempted to open up peer review started during the 1990s, see for instance the *Journal of Interactive Media in Education* (JIME) (Shum and Sumner 2001; Sumner amd Shum 1996) and the *British Medical Journal* (BMJ) (Smith 1999). The discussion on new models for peer review became intensified in recent years due to the results of a field trial at the journal *Nature* that

tested the use of this new form of review. In the PPR trial launched by *Nature*, authors whose manuscripts were selected for traditional closed peer review could opt to have their manuscripts posted to an open website for public comment: "Anyone can then respond to it by posting online comments, provided they are willing to sign them. Once *Nature*'s editors have received all the comments from their solicited confidential reviewers, the open website will cease to take comments, and all the opinions will be considered by the editors as well as the authors" (Anon 2006, p. 668). *Nature* rated the results of the trial as hardly encouraging. Very few authors opted to have their manuscripts hosted for public comment, and their manuscripts received very few comments. *Nature* concluded that most researchers are apparently "too busy, and lack sufficient career incentive, to venture onto a venue such as *Nature*'s website and post public, critical assessments of their peers' work" (Anon 2006, p. 972). And "although most authors found at least some value in the comments they received … editors did not think they contributed significantly to their decisions" (p. 972). However, in its trial *Nature* did not examine the reason why the public comments did not aid the editors in their decisions to accept or reject manuscripts for publication. How do these public comments differ from traditional reviewer comments? Why do the public comments add little to the assessment of the papers? It is possible that although the comments do not lead the editors to make more valid publication decisions than with reviewer comments alone, it could be that the suggestions for revising the manuscripts in the public comments can make a considerable contribution to improving the submissions.

This study examined the question as to the differences between the two groups of comments. Available to us for the investigation were the comments by reviewers and by interested members of the scientific community, conducted online using a comment sheet, for assessing manuscripts that had been submitted to *Atmospheric Chemistry & Physics* (ACP). We conducted a content analysis of these comments to determine differences in the aspects and issues that the reviewers focus on in their reviews.

Methods

Manuscript review at ACP

ACP was launched in September 2001. It is published by the European Geosciences Union (EGU; http://www.egu.eu) and Copernicus Publications (http://publications.copernicus.org). ACP is freely accessible via the Internet (http://www.atmos-chem-phys.org). According to Ulrich Pöschl, Chief Executive Editor of ACP, ACP has a two-stage publication process, including public peer review and interactive discussion (Pöschl 2004) that is described at the ACP website as follows: In the first stage, manuscripts that pass a rapid pre-screening process (access review) are immediately published as 'discussion papers' on the journal's *Atmospheric Chemistry and Physics Discussions* (ACPD) website: "The access review (pre-selection by the editor with optional advice from referees) is meant to avoid a potential overload of the discussion forum with papers that are clearly deficient or out of scope" (Pöschl 2004, p. 107).

The discussion papers are then made available for 'interactive public discussion,' during which the comments (reviewer comments, or RCs) of designated reviewers (these are usually the reviewers that already conducted the access review) (anonymous or attributed), additional comments by other interested members of the scientific community (short comments, or SCs) (attributed), and the authors' replies are published alongside the

discussion paper. "The interactive comments are published without peer review and revision, but can be censored in case of abusive commenting (personal offence, etc.). …To ensure publication precedence for authors and to provide a lasting record of scientific discussion, the discussion papers and interactive comments are permanently archived and fully citable" (Pöschl 2004, p. 107). According to Pöschl (2010) "the interactive peer review and public discussion offer direct feedback and public recognition for high-quality papers (authors' advantage); they prevent or minimize the opportunity for hidden obstruction and plagiarism (authors' advantage); they provide complete and citable documentation of critical comments, controversial arguments, scientific flaws and complementary information (referees' and readers' advantage); they reveal deficiencies and deter submissions of carelessly prepared manuscripts, thus helping to avoid/minimize the waste of time and effort for deficient submissions (referees', editors', publishers' and readers' advantage)" (p. 296).

Based on the revised manuscript (revised on the basis of RCs and SCs) and in view of the access peer review and interactive public discussion, the editor accepts or rejects the revised manuscript for publication in ACP. For this decision, further external reviewers may be asked to review the revision, if needed.

Database for this study

For the investigation of peer review at ACP we had data for 1,111 manuscripts that went through the complete ACP selection process in the years 2001–2006. These manuscripts reached one of the following final statuses: 958 (86 %) were published in ACPD and ACP, 74 (7 %) were published in ACPD but not in ACP (here, the editor rejected the revised manuscript), and 79 (7 %) were not published in either ACPD or ACP (these manuscripts were rejected during the access review). Some of the manuscripts ($n = 38$) submitted to ACP but not published there (because they were rejected during the access review, for example) were submitted by the authors to another journal and published there. According to Schultz (2010), there are two reasons for the high publication rate of submissions to ACP (see also Pöschl 2010): By using the public peer review and interactive discussion, (1) ACP can expect a high average quality of submitted manuscripts, and (2) ACP works harder than journals working with the traditional peer review to keep and improve the submissions.

Of the total 1,032 manuscripts that were published in ACPD, we included in the analysis only those manuscripts for which at least one SC was available (12 %). According to Elsevier: "Other than comments from invited reviewers, spontaneous comments from members of the scientific community have been relatively low" (House of Commons; Science and Technology Committee 2011, p. 27). In the content analysis of the RCs and SCs in this study, we were therefore able to include only a total of 390 comments, which were prepared for 119 manuscripts that were discussed interactively as ACPD papers. There are on average 3.3 comments available for each manuscript.

Category system used for the content analysis of the comments

Thematic analysis is the most common approach to content analysis (Krampen et al. 2007). However, only a few studies have been published up to now with thematic analyses of reviewer comments (e.g., Kumar et al. 2011). To prepare the thematic analysis of the comments in this study, we conducted a content analysis of 46 research studies that examined editors' and referees' criteria when assessing manuscripts and their reasons for

accepting or rejecting manuscripts. The results of this analysis are described in Bornmann et al. (2008). The goal of the analysis was to produce a catalogue as complete as possible of the different criteria and reasons used by editors and reviewers when assessing manuscripts. A total of 572 aspects and criteria from the 46 studies could be assigned to nine main areas: (1) Relevance, (2) Writing and Presentation, (3) Design and Conception, (4) Methods and Statistics, (5) Discussion of Results, (6) Reference to the Literature and Documentation, (7) Theory, (8) Author's Reputation and/or Institutional Affiliation, and (9) Ethics (Bornmann et al. 2008, pp. 418–419). This study used these nine main thematic areas, which are described in greater detail in the following, as a classification scheme for the content analysis of the SCs and RCs.

(1) The main area Relevance groups together reviewers' criteria and reasons concerning the future "gain" that could result from publication of a manuscript, with possible "gain" being scientific advancement, importance for journal readers, or the practical usefulness of the findings. These aspects have to do mainly with the importance, newness, and originality of the research study reported in the manuscript. Table 1 presents some examples of coding units from the ACP comments that were assigned to this thematic area or the other eight areas.

(2) The main area Writing and Presentation groups together mainly reviewers' criteria and reasons that refer to the formal qualities of a manuscript, such as writing style, written expression, spelling, grammar, and professional appearance of the manuscript (see here Shashok 2008). Also in this area are criteria and reasons with regard to following the journal's submission guidelines and appropriate length of the manuscript. Thoroughness also belongs here: Does the manuscript contain all of the necessary information in the different sections of the paper, presented completely and comprehensibly?

(3) The main area Design and Conception contains criteria and reasons referring to correct and logical conceptual framework and to the adequacy of the research design. Further aspects here are the internal consistency of a study, the plausibility of the research design with regard to the research question, the quality of sampling, the generalizability of the results, and replicability.

(4) The main area Methods and Statistics contains criteria and reasons that focus on the correctness, appropriateness, and newness of methods or statistical analyses. Also found here are criteria and reasons pertaining to the quality of the operationalization of key constructs and to the measurement of data.

(5) The main area Discussion of Results contains criteria and reasons pertaining mainly to whether the conclusions drawn in a manuscript are objective, correct, and properly based on the results. They also address the existence and clarity of a "take-home message" in the manuscript or the breadth or depth of the discussion section of a paper.

(6) The main area Reference to the Literature and Documentation contains criteria and reasons that focus on whether the research study reported in the manuscript is embedded in the framework of the relevant literature. The criteria and reasons pertain to the up-to-date review of the literature and the thoroughness of the author's review of the literature.

(7) The main area Theory contains criteria and reasons concerned with whether the manuscript contributes to theory development or whether the theory underlying the research study seems complete and sound.

Table 1 Examples of statements that a scientist wrote in a comment assessing a manuscript submitted to ACP and that could be assigned in this study to one of the nine main thematic areas considered by reviewers in manuscript assessment

Main area considered in manuscript assessment	Examples of statements (coding units)
Relevance	"The paper is relevant since it highlights several shortcomings in…"; "The paper provides an excellent platform for the discussion of…"; "The paper makes a stimulating contribution with respect to our knowledge of…"; "I find the subject original and of great interest"; "present new and substantial results"; "The results are important for the atmospheric science community"; "…is a very important contribution… and a major step towards…"; "This paper adds little to the existing body of literature…"; "I would like to applaud the authors for compiling a comprehensive and thorough overview and discussion of advances in the investigation and understanding of…"; "I have difficulties to see the relevance of the results…"
Writing and Presentation	"The paper is well written and it appears to be scientifically sound"; "…well structured and written"; "The introduction does not show a clear objective of this work"; "The authors used too many abbreviations and not all of them were properly handled"; "…I would favor a re-writing of this section…"
Design and Conception	"This paper is excellent from the point of view of sampling and analytical procedures"; "…analysis is thorough and generally well founded"; "The current records are too short"; "…the measurement methodology, techniques and instrumentation are all appropriate … as they appear well-executed"; "…a number of control experiments are required to eliminate the uncertainties…"; "… I appreciate author's approach of testing the model in details… before performing long simulations"
Methods and Statistics	"It is not demonstrated how well the parameterization works"; "…it is not appropriate to use… more sophisticated methods are needed"; "…one power-law is not sufficient to describe the overall behavior"; "it seems arbitrary to include … and not … in the calculation"; "…a more extensive sensitivity and uncertainty analysis must be carried out…"
Discussion of Results	"Interpretation of…is not straightforward and is not sufficient"; "The discussion of the results is very appropriate"; "The conclusions are consistent with the results, the authors never trying to go too far"; "…the speculation on possible mechanisms behind the observed features is very vague"; "The "Results" presented in next section are not sufficient to validate the method"
Reference to the Literature and Documentation	"Previous work is well reviewed"; "the authors should refer to this article…"; "… et al. didn't make any comments about…"; "…number and quality of references is appropriate…"; "…the authors do not want to go back to the earlier literature…"; "why are the results of…et al. and…et al. not included in the discussion?"
Theory	"I will say a few words about… which is incorrectly described in this …"; "The reason that the surface mode is thermodynamically favored is NOT… "
Author's Reputation and/or Institutional Affiliation	"Knowing the reliability and merits of the … group … I am also quite optimistic that fully transparent model documentation will be made available…"; "Building on earlier work … of previous modeling and analysis studies by others … he does a nice job in…"; "since … is co-author of this paper, I am surprised to see such a contradicting application of the cloud index introduced by him."

Table 1 continued

Main area considered in manuscript assessment	Examples of statements (coding units)
Ethics	"Half of the results have already been published elsewhere…"; "The authors intentionally avoid to cite literature showing…"; "…the authors try to promote their work trying to ignore what has been achieved in the last 30 years"

(8) Whereas main areas 1–7 above pertain to criteria used for assessing the content of manuscripts and the research reported in the manuscripts, there are also criteria used and reasons given by reviewers that refer to Author's Reputation and/or Institutional Affiliation. They focus on the scholarship demonstrated in the manuscript and the reputation of the authors in their research fields (see here also Kupfersmid 1988).

(9) The main area Ethics captures reviewers' criteria and reasons that focus on a manuscript's (or a research study's) compliance with scientific ethics or ethics standards within a discipline.

Conducting the content analysis and answering three questions

The content analysis of statements from the comments (RCs and SCs) was conducted by two members (coder 1 and coder 2) of our research team (they are the second and the third authors of this paper; they have doctorates in atmospheric physics), using the nine main areas following Bornmann et al. (2008). The two coders, working independently, assigned the coding units to the nine main areas. Following Gosden's (2003) method, the unit of analysis here was generally a complete sentence, although a coding unit may be longer than one sentence. Table 1 shows examples of statements that scientists wrote on a manuscript and that in this study could be assigned to one of the nine main thematic areas that the scientists focused on when assessing a manuscript. The two coders' assignments of the statements to the nine thematic areas were checked for disparities. All assignments for which disparities were determined were discussed within the team, and the team found a mutual solution for these assignments.

Along with assigning the statements to the nine thematic areas, the two coders also evaluated whether a statement judged the content of a manuscript as positive or negative. Table 2 shows an excerpt from the worksheet on which our research team recorded the positive and negative statements assigned to a thematic area (shown here: Relevance and Writing and Presentation). Positive statements were coded 1, and negative statements were coded −1. The table rows contain the code numbers assigned to a comment. The columns show the code assignments within a comment to the various thematic areas.

In addition to conducting the content analyses, the two coders gave each comment (SCs and RCs) a general rating with regard to the following three questions: (1) Does the comment help editors to decide whether a manuscript should be published? (5 rating categories, from 1 = very helpful to 5 = not helpful at all); (2) Does the comment help (authors) to improve the manuscript? (1 = very helpful, 5 = not helpful at all); (3) To what extent does the comment address the content of the manuscript? (1 = very helpful, 5 = not helpful at all). These questions aimed at identifying the functions that RCs and SCs generally fulfill (improvement versus selection) and at finding out how extensively the comments address the content of a manuscript.

Table 2 Excerpt from the worksheet that we used to record the positive (1) and negative (−1) statements assigned to a main thematic area (here: Relevance and Writing and Presentation)

Manuscript no.	Comment no.	Relevance			Writing and Presentation	
		First statement regarding this area	Second statement regarding this area	Third statement regarding this area	First statement regarding this area	Second statement regarding this area
1	1	1	1		1	−1
	2	−1	−1	−1	1	1
	3	1	1		−1	−1
2	1	1	−1	−1	1	1
	2	−1	1		−1	

In contrast to some studies on open peer review, we did not consider the quality of reviews in this study (van Rooyen et al. 1999; Walsh et al. 2000).

Statistical procedure

To test the extent to which there are statistically significant differences between the coders' ratings of the RCs and SCs on the three questions listed just above, we computed the Wilcoxon–Mann–Whitney test (StataCorp. 2011). To test agreement between the two coders' ratings on one and the same question for a comment, we used Cohen's Kappa coefficient (1960; StataCorp. 2011; von Eye and Mun 2005).

To identify differences between RCs and SCs with regard to the nine main areas considered in ACP manuscript assessment, we used multiple logistic regression models (Hosmer and Lemeshow 2000). These models are appropriate for the analysis of dichotomous (or binary) responses. Dichotomous responses arise when the outcome is the presence or absence of an event (Rabe-Hesketh and Everitt 2004). In this study, the binary response is coded 0 for a statement in a SC and 1 for a statement in a RC. With this—in view of the *Nature* trial—we aimed to find out the extent to which RCs and SCs differ in the aspects and issues of manuscripts that they assessed. In the model, the nine main areas considered in manuscript assessment are the independent variables, and the dichotomous responses (RC or SC) are the dependent variable.

The violation of the assumption of independent observations caused by including more than one comment for one single manuscript (see above) is considered in the logistic regression models by using the "cluster" option in Stata (2011). This option specifies that the comments are independent across the manuscripts but are not necessarily independent within one and the same manuscript (Hosmer and Lemeshow 2000, Sect. 8.3).

Results

General ratings of the comments with regard to the three questions

Table 3 shows the coders' answers to the three questions for a general rating of the RCs and SCs. The results in the table show how the coders rated the RCs and SCs on average and whether there are statistically significant differences between the ratings of RCs and SCs. Regarding the first two questions, concerning whether the comment helps

Table 3 The two coders' answers to three questions for a general rating of the reviewer comments (RC) and short comments (SC)

	n	Mean	SD	Median
(1) Does the comment help editors to decide whether a manuscript should be published? (5 rating categories, from 1 = very helpful to 5 = not helpful at all)				
Coder 1				
SC	142	3.74	1.14	4*
RC	247	2.49	0.95	2*
Coder 2				
SC	142	3.80	1.12	4*
RC	247	2.56	0.98	2*
Kappa Coder 1/Coder 2: .63[a]				
(2) Does the comment help to improve the submission? (1 = very helpful, 5 = not helpful at all)				
Coder 1				
SC	142	3.67	1.02	4*
RC	247	2.44	0.97	2*
Coder 2				
SC	142	3.55	1.03	4*
RC	247	2.34	1.06	2*
Kappa Coder 1/Coder 2: .45[a]				
(3) To what extent does the comment address the content of the manuscript? (1 = very clear, 5 = not clear at all)				
Coder 1				
SC	142	2.60	1.54	2*
RC	247	1.40	0.86	1*
Coder 2				
SC	142	2.42	1.41	2*
RC	247	1.37	0.85	1*
Kappa Coder 1/Coder 2: .69[a]				

Notes: * $p < 0.05$;

[a] Following guidelines by Fleiss (1981), the inter-coder agreement, with Kappa coefficients above 0.4, shows fair agreement (the higher the kappa value, the stronger the agreement; kappa = 1.0 is perfect agreement). According to Landis and Koch (1977), for questions 1 and 3, the inter-coder agreement is "substantial" as it lies between 0.61 and 0.80. For question 2, the inter-coder agreement is "moderate" as it lies between 0.41 and 0.60

(a) selection of manuscripts for publication, and (b) improvement of manuscripts, the two coders' ratings indicate that RCs (median = 2) are suitable for the two purposes and SCs (median = 4) are not. For the third question, regarding the extent to which the comments address the content of a manuscript, the results in Table 3 reveal that in the opinion of the two coders, both RCs and the SCs do so, with RCs doing so somewhat more than SCs do.

Assigning statements to the nine main areas

Table 4 indicates the number of statements that could be assigned to the nine main thematic areas considered by reviewers in their comments assessing ACP manuscripts. As the

Table 4 Number of statements from reviewer comments and short comments assigned to the main thematic areas (390 comments for 119 manuscripts) (arranged according to the sum of positive and negative statements over all comments, see the right-hand column)

Main area considered in manuscript assessment	Positive statements[a]			Negative statements[a]			Positive and negative statements		
	Minimal/maximal number in a comment	Average number per comment	Sum over all comments	Minimal/maximal number in a comment	Average number per comment	Sum over all comments	Minimal/maximal number in a comment	Average number per comment	Sum over all comments
Writing and Presentation	0/3	0.30	115	0/32	3.34	1304	0/32	3.64	1419
Discussion of Results	0/1	0.05	21	0/12	1.54	602	0/12	1.60	623
Methods and Statistics	0/1	0.08	3	0/11	1.12	435	0/11	1.19	465
Reference to the Literature and Documentation	0/2	0.03	11	0/7	0.65	255	0/7	0.68	266
Relevance	0/4	0.63	245	0/2	0.04	14	0/4	0.66	259
Design and Conception	0/2	0.09	33	0/3	0.10	40	0/3	0.19	73
Ethics	0/0	0.00	0	0/2	0.03	10	0/2	0.03	10
Theory	0/0	0.00	0	0/2	0.02	9	0/2	0.02	9
Author's Reputation and/or Institutional Affiliation	0/1	0.01	2	0/1	0.00	1	0/1	0.01	3
Total	0/7	1.17	457	0/52	6.85	2670	0/152	8.02	3127

Notes: [a]For positive statements, a scientist positively assessed an aspect which could be assigned (in the content analysis) to a main area (such as Relevance); negative statements relate to the critical assessment of a main thematic area by a scientist. The column under the heading "positive and negative statements" indicates the number of *all* coding units in the comments, i.e., no distinction is made between a statement that is worded positively or negatively

table shows, in the columns under the heading "Positive and negative statements," a total of 3,127 coding units in the 390 comments (for 119 manuscripts) were assigned to a main area (on average, 8.02 statements per comment). The majority of the statements ($n = 2{,}670$, on average, 6.85 statements per comment) criticize the content of a manuscript: only 457 statements refer favorably to a manuscript (on average, 1.17 positive comments per review). The fact that negative statements outweigh positive ones in ACP manuscript review is in line with the results of Bornmann et al. (2010), although all manuscripts that were included in that study were rejected by the editors for publication in AC-IE. However, the results of Bakanic et al. (1989) for the journal *American Sociological Review* also demonstrate that negative comments do not dominate only in the case of rejected manuscripts: "When addressing the authors of eventually published manuscripts, referees were more likely to preface the inevitable criticisms with positive comments. The ratio of negative to positive comments to the author was approximately 4:1 for accepted manuscripts and 5:1 for rejected manuscripts. This made interpreting the referees' comments more difficult for authors, since comments accompanying favorable recommendations were as critical as those with a recommendation to reject" (p. 643).

Most of the statements in Table 4 ($n = 1{,}419$; positive *and* negative) could be assigned to the thematic area Writing and Presentation. In addition, by way of comparison, many statements could be assigned to the thematic areas Discussion of Results ($n = 623$) and Methods and Statistics ($n = 65$). Comparing these results to the findings of Bornmann et al. (2010) study, which used the same category system with nine main thematic areas as this study did, it is noticeable that Relevance has clearly greater importance in traditional closed peer review at AC-IE than in the new PPR at ACP: In Bornmann et al. (2010) about one-third of the statements could be assigned to Relevance, but in this study it was only about one-tenth. Apparently, the possible "gain" of a manuscript in the form of scientific advancement, importance for journal readers, or the practical usefulness of the findings plays a greater role in peer review at AC than in peer review at ACP. This may be connected with the fact that AC-IE has a higher rejection rate (Bornmann and Daniel 2008) than ACP (Bornmann et al. 2010). If the rejection rate is higher, the importance of the accepted manuscripts (for journal readers, for the field, etc.) should be high.

The results of the logistic regression analyses

Two binary regression models were calculated with the number of positive and negative statements or only negative ones in a comment written on a manuscript submitted to ACP. The models are to determine the main areas of assessment that can be definitively associated with an RC or SC. In the regression models, six of the total nine thematic main areas could be considered. The three areas Ethics, Author's Reputation and/or Institutional Affiliation and Theory were not included, since for these areas the total number of the assigned coding units was too low (see the figures in Table 4).

As Table 5 shows, the first model (model A) uses the number of positive *and* negative statements and the second model (model B) uses the number of *only* negative statements. No model with the number of *only* positive statements was calculated, since in general, only a few positive statements can be found in the comments (see the figures in Table 4). Proportionate to the predominance of the negative statements in the comments, Table 5 shows that there are very similar results for regression models A and B: In both models, a statistically significant effect of the categories Writing and Presentation, Discussion of Results, and Relevance is demonstrated. The results of the regression models indicate that RCs have a higher number of statements in these three main areas than SCs do. For the

Table 5 Binary regression models calculating differences between short comments (=0) and reviewer comments (=1) assessing manuscripts submitted to ACP. Model A is based on the number of positive *and* negative statements; model B only applies to the number of *negative* statements

Main thematic area considered in manuscript assessment	Model A Positive and negative statements	Model B Only negative statements
Writing and Presentation	0.532*** (5.47)	0.549*** (5.71)
Discussion of Results	0.244* (2.18)	0.291** (2.98)
Methods and Statistics	0.178 (1.52)	0.190 (1.57)
Reference to the Literature and Documentation	−0.0480 (−0.29)	−0.0606 (−0.30)
Relevance	1.072*** (4.25)	1.266* (2.49)
Design and Conception	0.00141 (0.00)	−0.208 (−0.56)
Intercept	−1.790*** (−6.54)	−1.134*** (−6.05)
n_{reviews}	390	390
$n_{\text{manuscripts}}$	119	119

Notes: t statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

For a manuscript submitted to ACP, there are on average 3.3 comments. Therefore, standard errors are adjusted for this dependency in the data set

other four main areas, there are no significant differences between RCs and SCs in the number of statements in the comments.

Discussion

In this study, using the data from a comprehensive evaluation study on the public peer review process at ACP, we examined the extent to which RCs differ from SCs in this new, interactive form of peer review. To this purpose, we conducted a content analysis of RCs and SCs by interested members of the scientific community. For the content analysis of the comments, we used a category system with nine main thematic areas developed by Bornmann et al. (2008). The results of the analysis show that in contrast to interested members of the scientific community (in their SCs), reviewers (in their RCs) focus mainly on formal qualities of a manuscript, such as writing style (Writing and Presentation), the conclusions drawn in a manuscript (Discussion of Results), and the future "gain" that could result from publication of a manuscript (Relevance). In SCs, the lack of statements that can be assigned to these three main areas could be one of the reasons why *Nature*, in its trial of open peer review, found the comments of interested members of the scientific community to be of little help in selection decisions. To select manuscripts for publication in a high-impact journal, assessments of the relevance and interpretation of the results are particularly important. For the areas Methods and Statistics, Reference to the Literature and Documentation, and Design and Conception, this study found practically no differences between RCs and SCs.

The two coders' general ratings of the comments showed that the RCs supported the two main functions of peer review—namely, selection and improvement of publications—better than SCs did.

The results of this study thus complement observations on public peer review by Ulrich Pöschl, Chief Executive Editor of ACP, in recent years. Pöschl observed that "the

interactive comments published in *ACPD* comprise compliments and plaudits (e.g. *Atmos. Chem. Phys. Discuss.* 2003:3, S1107–8), constructive supplements (e.g. *Atmos. Chem. Phys. Discuss.* 2002:2, S530–2), and harsh criticism and controversy (e.g. *Atmos. Chem. Phys. Discuss.* 2003:3, S448–51 and S912–18). So far, however, no personal or abusive comments have occurred, and there has been no need for the editors to intervene" (Pöschl 2004, pp. 110–111). Mainly "discussion papers reporting controversial findings or innovations attract many interactive comments … non-controversial papers usually elicit comments only from the designated referees. Why would scientists invest effort and time commenting on papers which they find interesting but non controversial?" (Pöschl 2010, p. 300).

We would like to mention as a limitation of our study a point which refers to the temporal dimension in the publication process. Indeed, previous comments may have influenced reviewers and the general public. Let us take an individual willing to comment on a given paper. If comments are already present, he/she may first read them before posting his/her own comment, in order not to report issues already pointed out, for instance. Hence, the reading of past comments (1) may hinder the creation of new comments, or (2) may force to focus on points not addressed yet. We may believe that reviewers are the first ones to post comments, since they are asked to evaluate papers by the journal editors. Then, the subsequent readers from the general public may feel it unnecessary to repeat issues raised by the reviewers about formal quality, conclusions, and future gain. The general public may wish to address issues not covered by the reviewers so far, just to bring something new. In other words, we cannot completely eliminate that the difference between RCs and SCs are not due to (or influenced by) the order of comments.

Peer review, may it be closed, open, or even public, has attracted much debate in various research communities (see, e.g., Morrison 2006; Smith 1999; Wing and Chi 2011). This is still a hot topic (Wing and Chi 2011). In future studies, the order of comments should be considered by using, e.g., Markov models (Bornmann et al. 2008, 2009). Furthermore, it would be interesting in knowing to what extend the general public raised issues that were not identified by the reviewers.

Acknowledgments The research project, which is investigating quality assurance of interactive open access journals, is supported by a grant from the Max Planck Society (Munich, Germany). We thank Dr. Ulrich Pöschl, Chief Executive Editor of *Atmospheric Chemistry and Physics*, the Editorial Board of *Atmospheric Chemistry and Physics*, and Copernicus Publications (Katlenburg-Lindau, Germany) for permission to conduct the evaluation of the selection process of the journal, and we are grateful to the members of Copernicus Systems + Technology (Berlin, Germany) for their generous technical support during the carrying out of the study.

References

American Association for the Advancement of Science. (1998). *Report in AAAS/UNESCO/ICSU Workshop on Developing Practices and Standards for Electronic Publishing in Science*. Paris, October 12–14, 1998. Washington, DC: The American Association for the Advancement of Science.

Anon. (2006). Peer review on trial. *Nature, 441*(7094), 668.

Bakanic, V., McPhail, C., & Simon, R. J. (1989). Mixed messages: Referees' comments on the manuscripts they review. *Sociological Quarterly, 30*(4), 639–654.

Bedeian, A. G. (2004). Peer review and the social construction of knowledge in the management discipline. *Academy of Management Learning and Education, 3*(2), 198–216.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology, 45*, 199–245.

Bornmann, L., & 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*(38), 7173–7178. doi:10.1002/anie.200800513.

Bornmann, L., Marx, W., Schier, H., Thor, A., & Daniel, H.-D. (2010a). From black box to white box at open access journals: Predictive validity of manuscript reviewing and editorial decisions at *Atmospheric Chemistry and Physics*. *Research Evaluation, 19*(2), 105–118.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2008a). Latent Markov modeling applied to grant peer review. *Journal of Informetrics, 2*(3), 217–228.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2009). The influence of the applicants' gender on the modeling of a peer review process by using latent Markov models. *Scientometrics, 81*(2), 407–411.

Bornmann, L., Nast, I., & Daniel, H.-D. (2008b). Do editors and referees look for signs of scientific misconduct when reviewing manuscripts? A quantitative content analysis of studies that examined review criteria and reasons for accepting and rejecting manuscripts for publication. *Scientometrics, 77*(3), 415–432.

Bornmann, L., Weymuth, C., & Daniel, H.-D. (2010b). A content analysis of referees' comments: How do comments on manuscripts rejected by a high-impact journal and later published in either a low- or high-impact journal differ? *Scientometrics, 83*(2), 493–506.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Daniel, H.-D. (1993). *Guardians of science. Fairness and reliability of peer review*. Weinheim: Wiley–VCH.

Fleiss, J. (1981). *Statistical methods for rates and proportions*. New York, NY: Wiley–VCH.

Gosden, H. (2003). 'Why not give us the full story?': Functions of referees' comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes, 2*(2), 87–101.

Harnad, S. (1978). Inaugural editorial. *Behavioral and Brain Sciences, 1*(1).

Harnad, S. (2000). The invisible hand of peer review. *Exploit Interactive* (5).

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Chichester: Wiley.

House of Commons; Science and Technology Committee on alternative models for the federal funding of science. (2011). *Peer review in scientific publications. Eighth Report of Session 2010-12. Volume I: Report, together with formal minutes, oral and written evidence*. London: House of Commons.

Janeway, C. A. J. (1990). JMCI: The last issue. *Journal of Molecular and Cellular Immunology, 4*, 293.

Krampen, G., Becker, R., Wahner, U., & Montada, L. (2007). On the validity of citation counting in science evaluation: Content analyses of references and citations in psychological publications. *Scientometrics, 71*(2), 191–202.

Kumar, P., Rafiq, I., & Imam, B. (2011). Negotiation on the assessment of research articles with academic reviewers: Application of peer-review approach of teaching. *Higher Education, 62*(3), 315–332. doi:10.1007/s10734-010-9390-y.

Kupfersmid, J. (1988). Improving what is published: A model in search of an editor. *American Psychologist, 43*(8), 635–642.

Landis, J. R., & Koch, G. G. (1977). Measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

McCormack, N. (2009). Peer review and legal publishing: What law librarians need to know about open, single-blind, and double-blind reviewing. *Law Library Journal, 101*(1), 59–70.

Mizzaro, S. (2003). Quality control in scholarly publishing: Anew proposal. *Journal of the American Society for Information Science and Technology, 54*(11), 989–1005. doi:10.1002/Asi.10296.

Morrison, J. (2006). The case for open peer review. *Medical Education, 40*(9), 830–831. doi:10.1111/j.1365-2929.2006.02573.x.

Pöschl, U. (2004). Interactive journal concept for improved scientific publishing and quality assurance. *Learned Publishing, 17*(2), 105–113. doi:10.1087/095315104322958481.

Pöschl, U. (2010). Interactive open access publishing and peer review: The effectiveness and perspectives of transparency and self-regulation in scientific communication and evaluation. *Liber Quarterly, 19*(3/4), 293–314.

Rabe-Hesketh, S., & Everitt, B. (2004). *A handbook of statistical analyses using Stata*. Boca Raton, FL: Chapman & Hall/CRC.

Schultz, D. M. (2010). Rejection rates for journals publishing atmospheric science. *Bulletin of the American Meteorological Society, 91*(2), 231–243. doi:10.1175/2009BAMS2908.1.

Shashok, K. (2008). Content and communication: How can peer review provide helpful feedback about the writing? *BMC Medical Research Methods, 8*(3).

Shum, S. B., & Sumner, T. (2001). JIME: An interactive journal for interactive media. *First Monday, 6*(2–5).

Smith, R. (1999). Opening up BMJ peer review. *British Medical Journal, 318*(7175), 4–5. doi: 10.1136/bmj.318.7175.4.

StataCorp. (2011). *Stata statistical software: Release 12*. College Station, TX: Stata Corporation.

Sumner, T., & Shum, S. B. (1996). Open peer review & argumentation: Loosening the paper chains on journals. *ARIADNE, 5*.

van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: Arandomised trial. *British Medical Journal, 318*(7175), 23–27.

von Eye, A., & Mun, E. Y. (2005). *Analyzing rater agreement. Manifest variable methods*. Mahwah, NJ: Lawrence Erlbaum Associates.

Walsh, E., Rooney, M., Appleby, L., & Wilkinson, G. (2000). Open peer review: Arandomised controlled trial. *British Journal of Psychiatry, 176*, 47–51.

Wing, J. M., & Chi, E. H. (2011). Reviewing peer review. *Communications of the ACM, 54*(7), 10–11. doi: 10.1145/1965724.1965728.

Xin, H. (2006). Online sleuths challenge cell paper. *Science, 314*(5806), 1669. doi:10.1126/science.314.5806.1669a.