Is Interactive Open Access Publishing Able to Identify High-Impact Submissions? A Study on the Predictive Validity of *Atmospheric Chemistry and Physics* by Using Percentile Rank Classes

Lutz Bornmann
Max Planck Society, Office of Research Analysis and Foresight, Hofgartenstr. 8, D-80539 Munich,
E-mail: bornmann@gv.mpg.de

Hermann Schier and Werner Marx
Max Planck Institute for Solid State Research, Heisenbergstraße 1, D-70569 Stuttgart, Germany.
E-mail: {h.schier, w.marx}@fkf.mpg.de

Hans-Dieter Daniel
ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Zähringerstr. 24,
CH-8092 Zurich; University of Zurich, Evaluation Office, Mühlegasse 21, CH-8001 Zurich, Switzerland.
E-mail: daniel@evaluation.uzh.ch

In a comprehensive research project, we investigated the predictive validity of selection decisions and reviewers' ratings at the open access journal *Atmospheric Chemistry and Physics* (ACP). ACP is a high-impact journal publishing papers on the Earth's atmosphere and the underlying chemical and physical processes. Scientific journals have to deal with the following question concerning the predictive validity: Are in fact the "best" scientific works selected from the manuscripts submitted? In this study we examined whether selecting the "best" manuscripts means selecting papers that after publication show top citation performance as compared to other papers in this research area. First, we appraised the citation impact of later published manuscripts based on the percentile citedness rank classes of the population distribution (scaling in a specific subfield). Second, we analyzed the association between the decisions ($n = 677$ accepted or rejected, but published elsewhere manuscripts) or ratings (reviewers' ratings for $n = 315$ manuscripts), respectively, and the citation impact classes of the manuscripts. The results confirm the predictive validity of the ACP peer review system.

Introduction

The essential principle of journal peer review is that judgments about the scientific merit of a manuscript are made

by persons that have demonstrated competence to make such a judgment: the peers. Researchers submit manuscripts to a journal and their peers evaluate whether they should be published. In the light of peer review, an editorial decision is then made to publish or not publish in the journal (British Academy, 2007). According to Hansson (2006), peer review guarantees the quality of scientific knowledge products. Ideally, the peer review process prevents the publication of bad work (poorly conceived, designed, or executed studies) (Hames, 2007). "When the peer review process works, statements and opinions are not arbitrary, experiments and data meet certain standards, results follow logically from the data, merit rather than influence determines what is published, and researchers do not have to waste their time reading unfiltered material" (McCormack, 2009, p. 64).

In a survey of 3,040 academics on peer review conducted in 2007, 93% of the respondents disagreed that peer review is unnecessary (Publishing Research Consortium, 2008), but nonetheless there has been much criticism of the process over recent years. Critics of peer review charged the process with conservatism, as important and innovative papers are not recognized by reviewers (see here Hames, 2007). The process cannot guarantee "that 'good' science will prevail and 'bad' be rejected" (Geisler, 2000, p. 234). All in all, peer reviewers were said to do "a poor job of controlling quality" (Shatz, 2004, p. 2). Upon the background of these and similar criticisms of a process that is so central to science, it is considered necessary to examine the peer review process with the same

rigor that is commonly applied in scientific research generally (de Vries, Marschall, & Stein, 2009). These studies should aim to contribute to the peer review process being "carried out well and professionally" (Hames, 2007, p. 2).

Scientific journals that use peer review as a selection procedure have to deal with the following question concerning the predictive validity of the selection decisions: Are in fact the "best" scientific works selected from the manuscripts submitted? Reputable journals should additionally clarify whether selecting the "best" manuscripts also means selecting papers that after publication show top citation performance within their fields. According to our search of the literature, up to now only six empirical studies have been published on the level of predictive validity associated with editorial decisions. Research in this area is extremely labor intensive, since a validity test requires information regarding the fate of rejected manuscripts. All six studies were based exclusively on citation counts as a validity criterion. The editors of the *Journal of Clinical Investigation* (Wilson, 1978) and the *British Medical Journal* (Lock, 1985) undertook their own investigation into the question of predictive validity. Daniel (1993) and Bornmann and Daniel (2008a,b, 2010a) examined the editorial decisions at *Angewandte Chemie International Edition* (AC-IE). Opthof, Furstner, van Geer, and Coronel (2000) and McDonald, Cloft, and Kallmes (2009) looked at *Cardiovascular Research* and the *American Journal of Neuroradiology*, respectively. All of the studies showed that the editorial decisions (acceptance or rejection) for the different journals appear to reflect a rather high degree of predictive validity when citation counts are used as a validity criterion.

In a comprehensive research project, we investigated the quality of selection decisions and reviewers' ratings at the journal *Atmospheric Chemistry and Physics* (ACP). ACP is an open access (OA) journal, where the authors retain the copyright and the journal adopts the 'author/institution pays' policy (see here Giglia, 2007). Up to now, we published three publications from the project: (1) In Bornmann and Daniel (2010b) we examined the interrater reliability of ACP, i.e., "the extent to which two or more independent reviews of the same scientific document agree" (Cicchetti, 1991, p. 120). (2) Bornmann, Neuhaus, and Daniel (in press) investigated whether Thomson Reuters (Philadelphia, PA), for the Journal Citation Reports, Science Edition, correctly calculates the Journal Impact Factor (JIF) of ACP that publishes several versions of a manuscript within a two-stage publication process. (3) Bornmann, Marx, Schier, Thor, and Daniel (2010) examined the fate of manuscripts that were rejected by ACP, searched the JIFs of the journals in which rejected manuscripts were later published, and undertook a citation impact comparison of accepted and rejected but published elsewhere manuscripts.

As ACP is a high-impact journal in its field, in this study we examine whether selecting the "best" manuscripts among submitted also means selecting papers that after publication show top citation performance within their field. The first step of our approach in this study is to appraise the citation impact of manuscripts with different editorial decisions and reviewers' ratings based on the percentile citedness ranks of the population distribution (scaling in a specific subfield). The second step is to determine the association between the decisions and ratings, respectively, and the citation impact of the manuscripts.

With this evaluation of peer review at ACP we follow the recommendation by van Rooyen (2001) and Rowland (2002), among others, to conduct peer review research also for smaller or more specialist journals: "Our own studies have been conducted in a large general medical journal, and only in the UK. Other large studies have been conducted in the United States, but again in larger journals. It is important to know if the results obtained can be generalized to smaller or more specialist journals" (van Rooyen, 2001, p. 91). ACP is a smaller and a relatively new journal that publishes studies investigating the Earth's atmosphere and the underlying chemical and physical processes most relevant for research on global warming.

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 (www.atmos-chem-phys.org). It has the second highest annual JIF in the category "Meteorology & Atmospheric Sciences" (at 4.927 in the 2008 Journal Citation Reports). ACP has a two-stage publication process, with a "new" peer review process consisting of public peer review and interactive discussion (Koop & Pöschl, 2006; Pöschl, 2004) that is described on the ACP Website as follows: In the first stage, manuscripts that pass a rapid prescreening process (access review) are immediately published as "discussion papers" on the journal's Website (by doing this, they are published in *Atmospheric Chemistry and Physics Discussions*, ACPD). These discussion papers are then made available for "interactive public discussion," during which the comments of designated reviewers (usually, reviewers that already conducted the access review), additional comments by other interested members of the scientific community, and the authors' replies are published alongside the discussion paper.

During the discussion phase, the designated reviewers are asked to answer four questions according to the ACP's principal evaluation criteria (see http://www.atmospheric-chemistry-and-physics.net/review/ms_evaluation_criteria.html). The questions ask about scientific quality, scientific significance, presentation quality, and whether the manuscript is worthy of publication. With regard to scientific quality, for instance, the question is: "Are the scientific approach and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)?" The response categories for the question are: (1) excellent,

(2) good, (3) fair, and (4) poor. After the end of the discussion phase every author has the opportunity to submit a revised manuscript taking into account the reviewers' comments and the comments of interested members of the scientific community. Based on the revised manuscript 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 the Present 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 to 2006. These manuscripts reach 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). Of a total of 153 manuscripts, 38 (25%) that were submitted to ACP but not published by ACP were later submitted by the author to another journal and published there. No publication information was found for 115 (75%) of these manuscripts, whereby 70 of the 115 manuscripts (61%) were published in ACPD (Bornmann, Marx, et al., 2010). According to Schulz (2010), there are two reasons for the ACP's high publication rate of submissions (see also Pöschl, 2010): By using the public peer review and interactive discussion, (1) this journal can expect a high average quality of submitted manuscripts, and (2) it works harder than journals working with the traditional peer review to keep and improve the submissions.

In the examination of the predictive validity of acceptance versus rejection decisions in this study only 677 manuscripts could be included of the total of 1,066 manuscripts that were submitted to and later published in ACP or elsewhere. This reduction in the number of cases is mainly due to the fact that for many manuscripts no percentile citedness rank for evaluation of the citation impact could be found: There exists a field entry in Chemical Abstracts (CA) (Chemical Abstracts Services, Columbus, OH) in the literature database (see Reference Standard, below) for 698 manuscripts, and for 21 of those manuscripts there are no citation counts (see Conducting Citation Analysis, below). Due to the two reasons mentioned, the results of this study are valid mainly for manuscripts that were captured by CA—that is, chemistry and related sciences.

Reviewers' ratings on the scientific quality of the manuscripts were available for 552 (55%) of the 1,008 manuscripts that were reviewed in the discussion phase of ACP public review. This reduction in number is due to the fact that the publisher has stored the ratings electronically only since 2004. In the evaluation of predictive validity in this study we included ratings only for those manuscripts of the total of 552 manuscripts that were later published in ACP ($n = 496$). Through this restriction, factors were held constant in the evaluation that could have an undesired influence on the investigation of the association between reviewers' ratings and citation counts (for example, the prestige of the journal publishing the manuscript). As when examining the editorial decisions, here again only those manuscripts could be included in the analysis for which a percentile citedness rank class could be calculated for evaluation of the citation impact (see above). This resulted in a further reduction of the number of manuscripts from $n = 496$ to $n = 315$. Of these 315 manuscripts, 20% ($n = 62$) have one review, 61% ($n = 193$) have two, 16% ($n = 50$) have three, and 3% ($n = 10$) have four independent reviews. For the statistical analysis, for each manuscript the median of the independent ratings for the scientific quality was computed. According to Thorngate, Dawes, and Foddy (2009), the average error in ratings decreases with an increasing number of raters.

Conducting Citation Analysis

As there is currently no mathematical formula that can quantify the "quality" of an article (Figueredo, 2006), it is usual to examine the predictive validity of the peer review process using citation counts (van Raan, 2004). For Pendlebury (2008), "tracking citations and understanding their trends in context is a key to evaluating the impact and influence of research" (p. 3). For manuscripts published in ACP, ACPD, or elsewhere, we determined the number of citations for a fixed time window of 3 years including 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, Plume, McVeigh, Pringle, & Amin, 2007, p. 243). The citation analyses were conducted based on CA. CA is a comprehensive database of publicly disclosed research in chemistry and related sciences (see http://www.cas.org/).

CA does not include manuscripts published in ACPD as documents in the source index. But their citations are searchable using a method that is comparable to the 'Cited Reference Search' in Web of Science (WoS) (Thomson Reuters). For a manuscript the frequency of the various variants of the journal title of ACPD (for example, Atm Chem Phys Disc, Atm Chem Phys Discus, Atmos Chem Phys Disc) is searched in combination with the publication years within the totality of the references (citations) captured in the database and restricted to the correct time window. If a manuscript in our dataset was published not only in ACPD but also in another journal (mainly in ACP), the citation counts for both publications are added up. The addition of the two citation counts was conducted due to the fact that double count citations (that is, citation of both publications of a manuscript within one paper) occur only rarely (see here also Bloom, 2006).

Checking for double count citations was carried out using a recently developed routine for macro programming of the Messenger command language from STN International (Eggenstein-Leopoldshafen, Germany). This allowed

examination of the number of double count citations of the 958 individual ACP papers with the corresponding papers in ACPD in the Science Citation Index (SCI, Thomson Reuters) up to the present. Only 18 true double count citations were found where an ACP paper was cited together with the corresponding paper published in ACPD. In addition, we did a manual check of the number of double count citations for the complete ACP publication year 2004 as an example: For 2004 SCI shows 174 ACP papers as source items. The intersection of the 2,320 papers citing these ACP papers with the 371 papers citing the corresponding ACPD papers was 90 citing papers. In these 90 citing papers, at least one ACP paper from 2004 was cited together with an ACPD paper from 2004. A manual check of the citations of the ACP and ACPD papers in the citing papers revealed only three true double count citations. As using the two methods the citations across the complete time period were included, the number of double count citations for a 3-year window is smaller. Usually, the authors cited different papers and not corresponding ACP and ACPD papers.

Reference Standard

As the aim of the present study on the ACP was to evaluate a high-impact journal in the field of meteorology and atmospheric sciences, our focus was on peak performance. Our intention when conducting the citation analyses was not only to find out whether ACP peer review is able to select the "better" research (which we investigated in Bornmann, Marx, et al., 2010) but also to be able to identify "high-impact" submissions (see here Research Evaluation and Policy Project, 2005). Determining highly cited papers is possible only based on field-specific reference standards (Aksnes, 2003), as there are large differences in the expected citation frequencies between different (sub-)fields (Bornmann, Mutz, Neuhaus, & Daniel, 2008). For example, Radicchi, Fortunato, and Castellano, (2008) found that papers in biology (subject category of the journals where the papers appear; see Thomson Reuters) are cited on average 14.6 times, whereas papers in developmental biology are cited on average 38.67 times (see here also Bornmann & Daniel, 2009). Therefore, in this study the performance of manuscripts with acceptance or rejection decisions of ACP and reviewers' ratings was compared with international scientific reference standards. For this, Vinkler (1997) recommends calculating relative subfield citedness (RW) (see also van Raan, 1999): "Relative Subfield Citedness (Rw) (where W refers to 'world') relates the number of citations obtained by the set of papers evaluated to the number of citations received by a same number of papers published in journals dedicated to the respective discipline, field or subfield" (p. 164, see also Vinkler, 1986).

As Vinkler's (1997) definition of RW indicates, the determining of research fields in most studies of research evaluation is based on a classification of journals into subject categories developed by Thomson Reuters (Bornmann et al., 2008). "The Centre for Science and Technology Studies (CWTS) at Leiden University, the Information Science and Scientometrics Research Unit (ISSRU) at Budapest, and Thomson Scientific [now Thomson Reuters] itself use in their bibliometric analyses reference standards based on journal classification schemes" (Neuhaus & Daniel, 2009, p. 221). Each journal as a whole is classified as belonging to one or several subject categories. In general, this journal classification scheme proves to be of great value for research evaluation. But its limitations become obvious in the case of multidisciplinary journals such as *Nature* or *Science* (see, for example, Glänzel, Schubert, & Czerwon, 1999) and highly specialized fields of research (e.g., Glänzel, Thijs, Schubert, & Debackere, 2009; Kostoff, 2002; Schubert & Braun, 1996). Papers that appear in multidisciplinary journals cannot be assigned exclusively to one field, and for highly specialized fields no adequate reference values exist. To overcome the limitations of journal classification schemes, Neuhaus and Daniel (2009) proposed an alternative reference standard that is based on a paper-by-paper basis (see also Neuhaus & Daniel, 2010; Neuhaus, Marx, & Daniel, 2009). We follow that proposal in the present study. In contrast to a reference standard based on journal sets, where all papers in a journal are assigned to one and the same field, with the alternative reference standard every publication is associated with a single principal (sub-)field entry that makes clearly apparent the most important aspect of the work (see here also Kurtz & Henneken, 2007; Pendlebury, 2008).

For evaluation studies in chemistry and related fields, Neuhaus and Daniel (2009) proposed building reference values based on publication and citation data that refer to the subject areas of CA (see also van Leeuwen, 2007). For CA, CAS categorizes chemical publications into different subject areas (chemical fields, called "sections"). Every publication becomes associated with a single principal entry that makes clearly apparent the most important aspect of the work (Daniel, 1993). In contrast to the journal sets provided by Thomson Reuters, CA sections are assigned on a paper-by-paper basis (Bornmann, et al., 2008). According to Neuhaus and Daniel (2009), "the sections of *Chemical Abstracts* seem to be a promising basis for reference standards in chemistry and related fields for four reasons: (1) the wider coverage of the pertinent literature; (2) the quality of indexing; (3) the assignment of papers published in multidisciplinary and general journals to their respective fields; and (4) the resolution of fields on a more specific level (e.g., mammalian biochemistry) than in journal classification schemes (e.g., biochemistry and molecular biology). The proposed reference standard is transparent, reproducible and overcomes some limitations of the journal classification scheme of Thomson Scientific" (pp. 227–228).

For the present study, to set reference values we used publication and citation data for 25 CA subsections (e.g., "Electronic Spectroscopy") within 10 CA sections (e.g., "Optical, Electron, and Mass Spectroscopy and Other Related Properties") (see here Chemical Abstracts Service, 1997). For each of these 25 CA subsections we have in the sample at least one accepted or rejected (but published elsewhere)

manuscript. The reference standard for a section is based on the publications of the year, in which at least one manuscript in our dataset was published, and the citations of these publications for a fixed 3-year citation window. The manuscripts accepted or rejected (but published elsewhere) by ACP are mainly research articles and reviews. Because CAS "does not provide a distinct document type for research articles" (Neuhaus & Daniel, 2009, p. 226), the reference standards were generated by excluding publications with nonrelevant document types, such as conference proceedings.

Percentile Citedness Rank Procedure

In educational and psychological testing, percentile rank scores are used widely as a standard for comparison in order to judge a person's test scores (intelligence test scores, for example) based on a comparison with the percentiles of a calibrated sample (see Jackson, 1996). Percentile rank scores usually involve ranking the units, which in this study are papers, in ascending order according to a criterion, here citation counts (see Rousseau, 2005, for an example of the use of percentiles describing journal impact). Next, the frequencies with which papers with a certain citation count are found are accumulated successively across all papers (papers with citation count 0, 1, 2, . . .). The percentile rank score amounts to the fraction of the cumulative frequencies of the total number of all papers (Bornmann, Mutz, & Daniel, 2007).

Particularly in bibliometric analysis the use of percentile rank scores for evaluative purposes is very advantageous (see Evidence Ltd., 2007; Plomp, 1990), as no assumptions have to be made as to the distribution of citations; that is, the scores are applicable also for the (usually) right-skewed distributions of bibliometric data, distributed according to a power law (Adler, Ewing, Taylor, & Hall, 2009). In order to compare the citation rates of manuscripts, it is a common approach in bibliometrics to calculate arithmetic means. However, there are dangers in the orientation to the measures of central tendency: In the face of nonnormal distributed citation data, the arithmetic mean value can give a distorted picture of the kind of distribution (Bornmann et al., 2008), "and it is a rather crude statistic" (Joint Committee on Quantitative Assessment of Research, 2008, p. 2). Arithmetic means diagrams show mainly where publications with high citation rates are to be found. According to Evidence Ltd. (2007) "where bibliometric data must stand-alone, they should be treated as distributions and not as averages" (p. 10).

Through the use of percentile rank scores the scientist's papers can be assigned directly to unambiguous impact classes (through which the "experience-based" assignment to impact classes can be dropped, see here van Raan, 2004). The Essential Science Indicators (ESI) from Thomson Reuters, for example, offers baselines that provide expected citation rates for groups of papers in a specific field. The ESI percentiles table displays data on the minimum number of citations needed to meet the top 50%, 20%, 10%, 1%, 0.1%, and 0.01% of publications within 22 broad fields of science. For instance, in the "Chemistry" table, a value of 139 in the

1% column for the year 1999 (retrieved on January 15, 2010) indicates that the top 1% of papers in chemistry journals entered into the SCI in that year was cited at least 139 times.

Establishing the Reference Values With the Percentile Citedness Rank Procedure

Using the publication and citation data for each CA subsection, reference values were set in order to allow comparisons, on a common scale, of the citation counts of the manuscripts with acceptance or rejection decisions of ACP and reviewers' ratings and assigned by CAS to individual CA subsections (Bornmann, Mutz, Marx, Schier, & Daniel, conditionally accepted for publication). The reference values were computed using a percentile citedness rank procedure. First, the citation counts X_i received by the *i*th publications within *n* publications published in a given CA subsection were counted. Then the publications were ranked in increasing order

$$X_1 \leq X_2 \leq \ldots \leq X_n$$

where X_1 (X_n) denotes the number of citations received by the least (most) cited publication. Finally, in each CA subsection each individual publication was assigned a percentile citedness rank based on this distribution. If, for example, a single publication within a CA subsection had 50 citations, and this citation count was equal to or greater than the citation counts of 90% of all publications in the CA subsection, then the percentile citedness rank of this publication would be at best 90. The publication would be in the 90th percentile (or higher).

In the present study, within each CA subsection the percentiles were grouped in eight percentile citedness rank classes: (1) papers with a percentile less than the 25th percentile (called in the following: <25%); (2) papers from the 25th to the 50th percentile (25%); (3) papers from the 50th to the 75th percentile (50%); (4) papers from the 75th to the 90th percentile (75%); (5) papers from the 90th to the 95th percentile (90%); (6) papers from the 95th to the 99th percentile (95%); (7) papers from the 99th to the 99.9th percentile (99%); (8) top 0.1% of papers with a percentile equal to or greater than the 99.9th percentile. Based on the limit values (citation counts) of the percentile citedness rank classes that were set for each CA subsection in the present study, each individual accepted or rejected (but published elsewhere) manuscript was assigned to one of the eight percentile citedness rank classes.

The percentile citedness rank classes that we established for the individual CA subsections are very well suited for identifying the highly cited papers in a CA subsection. However, in evaluative bibliometrics there is uncertainty regarding what percentile citedness rank a paper should have to be considered highly cited. According to Tijssen, Visser, and van Leeuwen (2002) and Tijssen and van Leeuwen (2006), highly cited papers are those among the top 10% of the most highly cited papers—that is, papers in or greater than the 90th percentile (see also Lewison, Thornicroft,

Szmukler, & Tansella, 2007). In the ESI Thomson Reuters classifies as highly cited papers that belong to the top 1% of papers worldwide (papers in or greater than the 99th percentile), taking into account the field and year of publication (see also National Science Board, 2008).

Statistical Procedure

We tested the association between two categorial variables: (1) percentile citedness rank classes, and (2a) CA subsections or (2b) the categorized reviewers' median ratings, respectively, using Pearson's chi-square test (Agresti, 2002). A requirement for calculation of the asymptotic *p*-value (the term "asymptotic" means "given a sufficient sample size") based on Pearson's statistic is that for tables larger than 2×2 a minimum expected count of 1 is permissible as long as no more than about 20% of the cells have expected values below 5 (Cochran, 1954). As this requirement was not fulfilled by one of the tables in this study, we used the Monte Carlo option of StatXact (Cytel Software, 2010). The Monte Carlo option can generate an extremely accurate estimate of the exact *p*-value by the simple process of sampling tables from the reference set of all tables with the observed margins a large number of times (the default is 10,000 times). Provided each table is sampled in proportion to its hypergeometric probability, the fraction of sampled tables that are at least as extreme as the observed table gives an unbiased estimate of the exact *p*-value. That is, if *M* tables are sampled from the reference set, and *Q* of them are at least as extreme as the observed table, the Monte Carlo estimate of the exact *p*-value is:

$$\hat{P} = \frac{Q}{M}$$

Since the result of the statistical significance test is dependent on sample size and "statistical significance does not mean real life importance" (Conroy, 2002, p. 290), it is the strength of the association that is more interesting and important for interpreting the empirical finding. For calculating strength, we have to employ an additional measure of association, i.e., Cramer's V coefficient (Cramér, 1980). According to Kline (2004), Cramer's V "is probably the best known measure of association for contingency tables" (p. 151).

Results

Table 1 shows the number and proportions of all manuscripts ($n = 667$ accepted and $n = 10$ rejected but published elsewhere manuscripts) within the eight percentile citedness rank classes for different CA subsections. As the table shows, the 677 manuscripts were assigned by CAS to mainly two CA sections: "Air Pollution and Industrial Hygiene" ($n = 449$) and "Mineralogical and Geological Chemistry" ($n = 205$). Other sections ($n = 23$), such as "Water," are relatively seldom. In contrast to the journal AC-IE (a multidisciplinary high-impact chemistry journal), whose manuscripts are found distributed across approximately three-quarters of all CA sections (Bornmann &

Daniel, 2008b), the ACP thus has a clearly focused publication profile (or submission profile). In addition to the CA subsections Table 1 shows the citation limits for the 99th percentile, to give the reader an idea of the variety of the expected values for citations in the individual CA subsections. In order to belong to the top 1% of papers in a subsection, a manuscript needs between at least 11.7 (subsection "Optical, Electron, and Mass Spectroscopy and Other Related Properties – Moessbauer spectrometry") and at least 87.9 (subsection "Nuclear Phenomena – Cosmic rays and cosmology") citations. As the limits for the section "Optical, Electron, and Mass Spectroscopy and Other Related Properties" show, these clear citation differences are found not only between CA sections but also between the subsections of an individual CA section. This is a clear indication of the necessity to include reference values in the citation impact analysis.

The last three rows of Table 1 show the distribution of all accepted or rejected but published elsewhere manuscripts for the eight percentile citedness rank classes. Thirty manuscripts were assigned to the highest impact classes (99.9% and 99%). As the cumulative percents show, 4% of the manuscripts belong to the top 1% ($\geq$99th percentile) and 19% to the top 5% ($\geq$95th percentile) of papers within their CA subsection. Approximately one-third of the manuscripts are top 10% papers ($\geq$90th percentile) and can be considered to be highly cited papers (or the "elite set" of papers) according to the definition by Tijssen et al. (2002). In this connection, comparison values are available for the journal AC-IE: Results by Bornmann et al. (conditionally accepted for publication) show that between 9% ("Organometallic and Organometalloidal Compounds") and 26% ("Heterocyclic Compounds (One Hetero Atom)") of the manuscripts in their dataset (at AC-IE, accepted or rejected but published elsewhere manuscripts) are among the top 5% of papers in a CA section. Based on these figures, the impact of the ACP manuscripts can be considered to be extremely high, since the ACP figures refer to all manuscripts and not only manuscripts in the individual CA sections. This finding also illustrates very well that JIFs are not comparable across research fields: The JIF of AC-IE, at 10.879, is about twice as high as the JIF of ACP.

Table 2 (parts A and B) shows (A) the proportions of accepted, and (B) the numbers of rejected but published elsewhere manuscripts, within eight (A) and four (B) percentile citedness rank classes for different CA sections. As only 23 manuscripts could not be assigned to the CA sections "Air Pollution and Industrial Hygiene" and "Mineralogical and Geological Chemistry," these manuscripts were grouped together in the table (part A) in one category called "Other Section." As Table 2 (B) shows, 10 manuscripts in the dataset of this study were rejected by ACP. Whereas according to the results in Table 2 (B) no rejected manuscript belongs to the top 10% of the papers within a CA section (all manuscripts were assigned to a lower impact class), Table 2 (A) shows that about one-third of the accepted manuscripts belong to the top 10%. Even though only a small part of the already small number of manuscripts rejected by ACP could be included in the analysis of this study, this finding confirms the predictive

TABLE 1. Numbers and proportions (row percent) of accepted or rejected but published elsewhere manuscripts within eight percentile citedness rank classes for different CA subsections.

CA subsection (number of CA subsection)	Citation limit for 99%[a]	Percentile citedness rank class								
		<25%	25%	50%	75%	90%	95%	99%	99.9%	Total
Mineralogical and Geological Chemistry (53)										
Reviews (00)	39.0				1			1	1	3
Igneous rocks (03)	22.0	2								2
Cosmochemistry and meteorites (09)	54.4	1	2			1				4
Atmosphere (10)	28.0	6	22	53	45	32	27	7	1	193
Geochemistry of water (11)	15.0			2						2
Other (12)	34.3					1				1
Total		9	24	55	46	34	27	8	2	205
Row percent		4	12	27	22	17	13	4	1	100
Row percent (cumulative)		100	96	84	57	35	18	5		
Air Pollution and Industrial Hygiene (59)										
Reviews (00)	85.5	2	1	3	5	1	2	1	1	16
Analysis (01)	22.2	2	5	10	11	8	4	1	1	42
Air pollutants and air pollution (02)	30.0	19	71	68	88	61	60	14	2	383
Combustion engine exhaust gas and catalytic converters (03)	26.9	1		3	2	1				7
Industrial waste gases (04)	21.0				1					1
Total		24	77	84	107	71	66	16	4	449
Row percent		5	17	19	24	16	15	4	1	100
Row percent (cumulative)		100	95	78	59	35	19	4[b]		
Water (61)										
Source (01)	16.9	1		2	2	1	1			7
Water pollution (02)	23.0						1			1
Analysis (03)	34.7		2							2
Atmospheric precipitation (09)	12.6			1	1					2
Optical, Electron, and Mass Spectroscopy and Other Related Properties (73)										
Electronic spectroscopy (04)	38.3			1						1
Moessbauer spectrometry (07)	11.7			1						1
Astrophysical spectra (09)	85.6	2								2
Spectrometers and optical apparatus (11)	41.0				1					1
Others										
Food and Feed Chemistry – Analysis (17-01)	23.0					1				1
Surface Chemistry and Colloids – Liquid-gas systems (66-01)	22.0						1			1
Nuclear Phenomena – Cosmic rays and cosmology (70-07)	87.9	1								1
Radiation Chemistry, Photochemistry, and Photographic and Other Reprographic Processes – Radiation chemistry and photochemistry (74-01)	37.0					1				1
Inorganic Analytical Chemistry – Reagents (79-03)	46.4		1							1
Terpenes and Terpenoids – Monoterpenes (C10), including cannabinoids, chrysanthemic acids, and iridoid aglycons (30-10)	33.8					1				1
Total (all subsections)		37	104	144	157	109	96	24	6	677
Row percent		5	15	21	23	16	14	4	1	100
Row percent (cumulative)		100	95	79	58	35	19	4[b]		

Note. Citation impact of 677 accepted or rejected, but published elsewhere manuscripts for a fixed time window of 3 years including the publication year.
[a] The values indicate the limit in order to belong to the top 1% of papers (that is, to be in die 99th percentile citedness rank class or higher) in a CA subsection. Depending on the subsection the values refer to different publication years of the papers (from 2002 to 2006) but always based on a three-year citation window.
[b] Due to a round-off error, the resulting value is 4 and not 5.

validity of the ACP selection decisions as already ascertained by Bornmann, Marx, et al. (2010). The nonsignificant result of the chi-square test and the small effect size for Table 2 (A) in addition show that the successful selection of the manuscripts by the editors applies to all CA sections represented in the dataset: The percentages for the assignment of the manuscripts to the individual impact classes in the table hardly differ between the individual sections in the table.

TABLE 2. (A) Proportions of accepted and (B) numbers of rejected but published elsewhere manuscripts within eight (A) and four (B) percentile citedness rank classes for different CA sections.

(A) Accepted manuscripts (proportions)

	CA section (number of CA subsection) (number of manuscripts within the CA section)				
Percentile citedness rank class	Air Pollution and Industrial Hygiene (59) (n = 441)	Mineralogical and Geological Chemistry (53) (n = 203)	Other section (n = 23)	Total (n = 667)	Cumulative percent
<25%	5	3	17	5	100
25%	17	12	13	15	95
50%	18	27	22	21	80
75%	24	23	18	23	59
90%	16	17	17	16	36
95%	15	13	13	15	20
99%	4	4		4	5
99.9%	1	1		1	
Total	100	100	100	100	

(B) Rejected but published elsewhere manuscripts (numbers)

	CA section (number of CA subsection)		
Percentile citedness rank class	Air Pollution and Industrial Hygiene (59)	Mineralogical and Geological Chemistry (53)	Total
<25%	2	2	4
25%	2	0	2
50%	2	0	2
75%	2	0	2
Total	8	2	10

Note. $\chi^2_{14} = 17.3$, $P = .24$, Cramer's V = 0.11.
Citation impact of 667 accepted manuscripts for a fixed time window of 3 years including the publication year.
Citation impact of 10 rejected, but published elsewhere manuscripts for a fixed time window of three years including the publication year.

Because almost all manuscripts that were included in the above citation analysis for accepted or rejected but published elsewhere manuscripts belong to one single group (published in ACP and ACPD), we undertook a differentiation within this group based on the reviewers' ratings (median reviewers' ratings on the scientific quality of the manuscripts) for further investigation of the predictive validity of the ACP peer review process. In this analysis we examined the extent to which the ratings for accepted manuscripts received during the discussion phase of the peer review process correspond with the assignment to the percentile citedness rank classes. Since this analysis entailed a reduction of the number of cases to 315 manuscripts (see Database for the Present Study, above), categories of the variables "reviewers' median ratings" and "percentile citedness rank classes" that had few cases were grouped together in one category (see notes to Table 3).

Table 3 shows the citation impact for the manuscripts with different median ratings. As the results of the chi-square test show, there is no statistically significant correspondence between citation impact and ratings. The Cramer's V value indicates a small effect size. As our results in Bornmann, Marx, et al. (2010) show, the result of the statistical significance test reported in Table 3 does not tally with the finding that we obtained with the median citation counts of manuscripts having different reviewers' ratings. That analysis showed statistically significant differences: A better median rating was associated with a statistically significant increase in the median citation count. However, about 200 additional manuscripts could be included in the analysis by Bornmann, Marx, et al. (2010), and as we mentioned in the Methods section, the statistical significance test is dependent on sample size. Despite the fact that the result of the chi-square test in Table 3 is not statistically significant, the proportions in the table still show a tendency in the expected direction: Whereas, for example, 29% of the manuscripts with a median rating ≤ 1.5 belong to the top 5% of cited papers in their field ($\geq 95\%$), this is 25% and 18% of the manuscripts with a median rating of 2 and ≥ 2.5.

Discussion

The strongest reservation about OA journals like ACP is doubt as to whether they achieve sufficient quality control (Joint Information Systems Committee, 2004). "In the open-access business model, it is widely accepted that authors (or their funding agencies or universities) pay. This means that ... the earnings of the journal are directly dependent on

TABLE 3. Citation impact (percentile citedness rank class) of accepted manuscripts with different reviewers' median ratings.

Percentile citedness rank class[b]	Reviewers' median rating[a]			Total ($n = 315$)
	≤ 1.5 (n = 62)	2 (n = 152)	≥ 2.5 (n = 101)	
$\leq 25\%$	7	17	16	15
50%	16	16	29	21
75%	29	22	21	23
90%	19	20	16	18
$\geq 95\%$	29	25	18	23
Total	100	100	100	100

Note. $\chi^2_8 = 13.7$, $p = .09$, Cramer's V $= 0.15$
Citation impact of 315 accepted manuscripts for a fixed time window of 3 years including the publication year.
[a]1 = excellent, 2 = good, 3 = fair, 4 = poor. The manuscripts with the median ratings 1 ($n = 28$) and 1.5 ($n = 34$) were grouped in the category ≤ 1.5, and the manuscripts with the median ratings 2.5 ($n = 55$), 3 ($n = 36$), 3.5 ($n = 5$), and 4 ($n = 5$) were grouped in the category ≥ 2.5.
[b]The classes <25% ($n = 9$) and 25% ($n = 37$) were grouped in the category $\leq 25\%$, and the classes 95% ($n = 57$), 99% ($n = 13$), and 99.9% ($n = 4$) were grouped in the category $\geq 95\%$.

the number of articles published. Only fools believe that editors wouldn't then tend towards acceptance of a manuscript in the many borderline cases" (Gölitz, 2010, p. 4). According to Taylor, Perakakis, and Trachana (2008), "one may argue that editors of OA ... journals, pressured by their commercial employers, may be forced to accept articles of lower quality in order to increase the number published and therefore the journal's income" (p. 31). As was discovered recently, Merck—a global pharmaceutical and chemical company (see http://www.merck.de/)—"paid an undisclosed sum to Elsevier [Amsterdam, the Netherlands] to produce several volumes of a publication that had the look of a peer-reviewed medical journal, but contained only reprinted or summarized articles—most of which presented data favorable to Merck products—that appeared to act solely as marketing tools with no disclosure of company sponsorship" (Grant, 2009).

In this study we evaluated the quality of the manuscript selection process at an OA journal, taking ACP as an example. We investigated whether the ACP peer review system in fact selects papers that after publication show top citation performance within their fields. As the results concerning different median reviewers' ratings show, there is no statistically significant correspondence between citation impact and ratings. However, the proportions in the table still show a tendency in the expected direction. Furthermore, the results reveal that about one-third of the ACP papers in recent years belong to the top 10% of papers in their fields. With about one-third, more papers are in the top 10% than one would expect for randomly chosen papers. If papers are chosen randomly from the underlying population (here: all publications within the relevant CA sections), one would expect this value to be 10% (see here Sheskin, 2007). Compared to the results found for the journal AC-IE (Bornmann et al., conditionally accepted for publication), this finding indicates high predictive validity of the ACP selection decisions. This not only fundamentally confirms the findings of other studies on the predictive validity of closed peer review (see Introduction) but also the first results on the predictive validity of the selection process at ACP that we presented in Bornmann, Marx, et al. (in press). In this

evaluation study we compared the median citation impact of papers that were accepted for publication in ACP after a positive review or rejected for publication in ACP after a negative review but submitted and published elsewhere. We found clear citation impact differences between the manuscript groups.

All in all, our results on the predictive validity of the ACP peer review system can support the high expectations that Pöschl (2010), chief executive editor of ACP, has of the new selection process at the journal: "The two-stage publication process stimulates scientists to prove their competence via individual high-quality papers and their discussion, rather than just by pushing as many papers as possible through journals with closed peer review and no direct public feedback and recognition for their work. Authors have a much stronger incentive to maximize the quality of their manuscripts prior to submission for peer review and publication, since experimental weaknesses, erroneous interpretations, and relevant but unreferenced earlier studies are more likely to be detected and pointed out in the course of interactive peer review and discussion open to the public and all colleagues with related research interests. Moreover, the transparent review process prevents authors from abusing the peer review process by delegating some of their own tasks and responsibilities to the referees during review and revision behind the scenes" (p. 297).

Acknowledgments

References

Adler, R., Ewing, J., Taylor, P., & Hall, P.G. (2009). A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). Statistical Science, 24(1), 1–28.

Agresti, A. (2002). Categorical data analysis. Hoboken, NJ: John Wiley & Sons.

Aksnes, D.W. (2003). Characteristics of highly cited papers. Research Evaluation, 12(3), 159–170.

Bloom, T. (2006). Systems: Online frontiers of the peer-reviewed literature. The Internet is allowing much more interactive science publishing. Retrieved July 6, 2006, from http://www.nature.com/nature/peerreview/debate/nature05030.html

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). 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. (2009). Universality of citation distributions. A validation of Radicchi et al.'s relative indicator $c_f = c/c_0$ at the micro level using data from chemistry. Journal of the American Society for Information Science and Technology, 60(8), 1664–1670.

Bornmann, L., & Daniel, H.-D. (2010a). Citation speed as a measure to predict the attention an article receives: An investigation of the validity of editorial decisions at Angewandte Chemie International Edition. Journal of Informetrics, 4(1), 83–88.

Bornmann, L., & Daniel, H.-D. (2010b). Reliability of reviewers' ratings at an interactive open access journal using public peer review: A case study on Atmospheric Chemistry and Physics. Learned Publishing, 23(2), 124–131.

Bornmann, L., Marx, W., Schier, H., Thor, A., & Daniel, H.-D. (2010). 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. (2007). The b-index as a measure of scientific excellence. A promising supplement to the h index. Cybermetrics, 11(1), paper 6.

Bornmann, L., Mutz, R., Marx, W., Schier, H., & Daniel, H.-D. (conditionally accepted for publication). 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? Journal of the Royal Statistical Society.

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.

Bornmann, L., Neuhaus, C., & Daniel, H.-D. (in press). The effect of a two-stage publication process on the Journal Impact Factor: A case study on the interactive open access journal Atmospheric Chemistry and Physics. Scientometrics.

British Academy. (2007). Peer review: The challenges for the humanities and social sciences. London: British Academy.

Chemical Abstracts Service. (1997). Subject coverage and arrangement of abstracts by sections in Chemical Abstracts. Columbus, OH: Chemical Abstracts Service (CAS).

Cicchetti, D.V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. Behavioral and Brain Sciences, 14(1), 119–135.

Cochran, W.G. (1954). Some methods for strengthening the common X^2 tests. Biometrics, 10(4), 417–451.

Conroy, R.M. (2002). Choosing an appropriate real-life measure of effect size: The case of a continuous predictor and a binary outcome. The Stata Journal, 2(3), 290–295.

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.

Cramér, H. (1980). Mathematical methods of statistics. Princeton, NJ: Princeton University Press.

Cytel Software Corporation. (2010). StatXact: Version 9. Cambridge, MA: Cytel Software Corporation.

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

de Vries, D.R., Marschall, E.A., & Stein, R.A. (2009). Exploring the peer review process: What is it, does it work, and can it be improved? Fisheries, 34(6), 270–279.

Evidence Ltd. (2007). The use of bibliometrics to measure research quality in UK higher education institutions. London: Universities UK.

Figueredo, E. (2006). The numerical equivalence between the impact factor of journals and the quality of the articles. Journal of the American Society for Information Science and Technology, 57(11), 1561.

Geisler, E. (2000). The metrics of science and technology. Westport, CT: Quorum Books.

Giglia, E. (2007). Open access in the biomedical field: A unique opportunity for researchers (and research itself). Europa Medicophyisica, 43(2), 203–213.

Glänzel, W., Schubert, A., & Czerwon, H.J. (1999). An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. Scientometrics, 44(3), 427–439.

Glänzel, W., Thijs, B., Schubert, A., & Debackere, K. (2009). Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. Scientometrics, 78(1), 165–188.

Gölitz, P. (2010). Twitter, Facebook, and Open Access. Angewandte Chemie International Edition, 49(1), 4–6.

Grant, B. (2009). Merck published fake journal. Retrieved March 25, 2010, from http://www.the-scientist.com/blog/display/55671/

Hames, I. (2007). Peer review and manuscript management of scientific journals: Guidelines for good practice. Oxford, United Kingdom: Blackwell.

Hansson, F. (2006). Research evaluation and organisational learning in the university: A possible coexistence? International Journal of Learning and Change, 1(3), 267–284.

Jackson, C. (1996). Understanding psychological testing. Leicester, United Kingdom: British Psychological Society.

Joint Committee on Quantitative Assessment of Research. (2008). Citation statistics. A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). Berlin, Germany: International Mathematical Union (IMU).

Joint Information Systems Committee. (2004). Journal authors survey report. Truro, United Kingdom: Key Perspectives Ltd.

Kline, R.B. (2004). Beyond significance testing: Reforming data analysis methods in behavioral research. Washington, DC: American Psychological Association.

Koop, T., & Pöschl, U. (2006). Systems: An open, two-stage peer-review journal. The editors of Atmospheric Chemistry and Physics explain their journal's approach. Retrieved 26 June 2006, from http://www.nature.com/nature/peerreview/debate/nature04988.html

Kostoff, R.N. (2002). Citation analysis of research performer quality. Scientometrics, 53(1), 49–71.

Kurtz, M.J., & Henneken, E.A. (2007). Open Access does not increase citations for research articles from The Astrophysical Journal. Retrieved September 10, 2007, from http://arxiv.org/abs/0709.0896

Lewison, G., Thornicroft, G., Szmukler, G., & Tansella, M. (2007). Fair assessment of the merits of psychiatric research. British Journal of Psychiatry, 190, 314–318.

Lock, S. (1985). A difficult balance: Editorial peer review in medicine. Philadelphia: ISI Press.

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

McDonald, R.J., Cloft, H.J., & Kallmes, D.F. (2009). Fate of manuscripts previously rejected by the American Journal of Neuroradiology: A follow-up analysis. American Journal of Neuroradiology, 30(2), 253–256.

National Science Board. (2008). Science and Engineering Indicators 2008 (NSB 08-01A) (Vol. 2). Arlington, VA: National Science Foundation (NSF).

Neuhaus, C., & Daniel, H.-D. (2009). A new reference standard for citation analysis in chemistry and related fields based on the sections of Chemical Abstracts. Scientometrics, 78(2), 219–229.

Neuhaus, C., & Daniel, H.D. (2010). How good is research? Chemistry World, 7(3), 42.

Neuhaus, C., Marx, W., & 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(1), 176–183.

Opthof, T., Furstner, F., van Geer, M., & Coronel, R. (2000). Regrets or no regrets? No regrets! The fate of rejected manuscripts. Cardiovascular Research, 45(1), 255–258.

Pendlebury, D.A. (2008). Using bibliometrics in evaluating research. Philadelphia: Research Department, Thomson Scientific.

Plomp, R. (1990). The significance of the number of highly cited papers as an indicator of scientific prolificacy. Scientometrics, 19(3–4), 185–197.

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

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.

Publishing Research Consortium. (2008). Peer review in scholarly journals: perspective of the scholarly community—an international study. Bristol, United Kingdom: Publishing Research Consortium.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. Proceedings of the National Academy of Sciences of the United States of America, 105(45), 17268–17272.

Research Evaluation and Policy Project. (2005). Quantitative indicators for research assessment – A literature review (REPP discussion paper 05/1). Canberra, Australia: Research Evaluation and Policy Project, Research School of Social Sciences, The Australian National University.

Rousseau, R. (2005). Median and percentile impact factors: A set of new indicators. Scientometrics, 63(3), 431–441.

Rowland, F. (2002). The peer-review process. Learned Publishing, 15(4), 247–258.

Schubert, A., & Braun, T. (1996). Cross-field normalization of scientometric indicators. Scientometrics, 36(3), 311–324.

Schultz, D.M. (2010). Rejection rates for journals publishing atmospheric science. Bulletin of the American Meteorological Society, 91(2), 231–243.

Shatz, D. (2004). Peer review: A critical inquiry. Lanham, MD: Rowman & Littlefield.

Sheskin, D. (2007). Handbook of parametric and nonparametric statistical procedures (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.

Taylor, M., Perakakis, P., & Trachana, V. (2008). The siege of science. Ethics in Science and Environmental Politics, 8, 17–40.

Thorngate, W., Dawes, R.M., & Foddy, M. (2009). Judging merit. New York: Psychology Press.

Tijssen, R., & van Leeuwen, T. (2006). Centres of research excellence and science indicators. Can 'excellence' be captured in numbers? In W. Glänzel (Ed.), Ninth International Conference on Science and Technology Indicators (pp. 146–147). Leuven, Belgium: Katholieke Universiteit Leuven.

Tijssen, R., Visser, M., & van Leeuwen, T. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? Scientometrics, 54(3), 381–397.

van Leeuwen, T.N. (2007). Modelling of bibliometric approaches and importance of output verification in research performance assessment. Research Evaluation, 16(2), 93–105.

van Raan, A.F.J. (1999). Advanced bibliometric methods for the evaluation of universities. Scientometrics, 45(3), 417–423.

van Raan, A.F.J. (2004). Measuring science. Capita selecta of current main issues. In H.F. Moed, W. Glänzel, & U. Schmoch (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.

van Rooyen, S. (2001). The evaluation of peer-review quality. Learned Publishing, 14(2), 85–91.

Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. Scientometrics, 10(3–4), 157–177.

Vinkler, P. (1997). Relations of relative scientometric impact indicators. The relative publication strategy index. Scientometrics, 40(1), 163–169.

Wilson, J.D. (1978). Peer review and publication. Journal of Clinical Investigation, 61(4), 1697–1701.