Scientometrics (2011) 89:835–843
DOI 10.1007/s11192-011-0472-0

Does the *h* index for assessing single publications really work? A case study on papers published in chemistry

Lutz Bornmann · Hermann Schier · Werner Marx · Hans-Dieter Daniel

Received: 16 June 2011 / Published online: 6 August 2011

Abstract Schubert (Scientometrics, 78:559–565, 2009) showed that "a Hirsch-type index can be used for assessing single highly cited publications by calculating the *h* index of the set of papers citing the work in question" (p. 559). To demonstrate that this single publication *h* index is a useful yardstick to compare the quality of different publications; the index should be strongly related to the assessment by peers. In a comprehensive research project we investigated the peer review process of the *Angewandte Chemie International Edition*. The data set contains manuscripts reviewed in the year 2000 and accepted by the journal or rejected but published elsewhere. Single publication *h* index values were calculated for a total of 1,814 manuscripts. The results show a correlation in the expected direction between peer assessments and single publication *h* index values: After publication, manuscripts with positive ratings by the journal's reviewers show on average higher *h* index values than manuscripts with negative ratings by reviewers (and later published elsewhere). However, our findings do not support Schubert's (2009) assumption that the additional dimension of indirect citation influence contributes to a more refined picture of the most cited papers.

Keywords Single publication *h* index · Journal peer review · Chemistry · *h* index

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

H. Schier · W. Marx
Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany

H.-D. Daniel
ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Zähringerstr 24, 8092 Zurich, Switzerland

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

Introduction

Today, the *h* index (Hirsch 2005) is a widely used measure of scientific performance: "The automatic calculation of *h*-indices has even become a built-in feature of major bibliographic databases such as Web of Science and Scopus" (van Eck and Waltman 2008, p. 263). The *h* index is usually used as a single number providing information on the size of the publication output of a scientist or journal and the citation impact of these publications (Bornmann and Daniel 2007; Bornmann and Daniel 2009c). Recently, Schubert (2009) proposed calculating the *h* index for a single publication: In a first step, a list is drawn up of papers citing the publication. In a second step, the number of times the citing publications were cited is determined. The *h* index for the single publication is the number for which the rank number of a publication on the list of citing publications sorted by times cited matches the number of citations received by this publication.

Schubert (2009) justified his proposal to calculate an *h* index for single publications as follows:

> Citation indicators usually measure the 'direct impact' of publications, i.e., the amount of the citations received (whether in the form of simple counts, weighted sums or normalized units). Undoubtedly, however, publications may exert influence also indirectly, e.g., through their presence in reference lists … It seems therefore reasonable to construct indicators that take into account not only the direct [but] also the indirect citation influence of publications. (p. 560)

Precisely for highly cited publications the additional consideration of indirect citation influence is said to lead to a more refined picture of the performance of a single publication than with citation counts only.

To demonstrate that the single publication *h* index is a useful yardstick to compare different publications the index should be strongly related to the assessment by peers (Cole 1989). Peer review is the principal mechanism for quality control in science publishing. But it is not perfect. Peers are not prophets, but ordinary human beings with their own opinions, strengths, and weaknesses (Bornmann 2011). Since it is yet the best available mechanism, it is used in this study to examine the single publication *h* index of manuscripts. In a comprehensive research project we investigated the peer review process of the *Angewandte Chemie International Edition* (AC-IE). Based on our AC-IE database we examined to what extent the single publication *h* index of manuscripts submitted to AC-IE and published in the journal or rejected by the journal and published elsewhere is systematically associated (1) with the publication decision of the AC-IE editors (accepted or rejected but published elsewhere), and (2) with external reviewers' assessment of the importance of a manuscript.

Methods

Manuscript review at AC-IE

A manuscript submitted to AC-IE usually undergoes internal and external review. First, editors at the journal evaluate whether the manuscript contributes to the development of an important area of research (internal review). If the editorial office finds that this is so, the submitted manuscript is sent to several independent reviewers (external review), who review it using an evaluation form and a comment sheet (Bornmann et al. 2010).

The evaluation form for reviewers contains, among other things, the following question about the importance of a manuscript: (1) "How important do you consider the results?" (four response categories: very important, important, less important, unimportant). The journal editors make the decision to accept or reject a manuscript for publication on the basis of these reviews and their own evaluations (Bornmann and Daniel 2009b, 2010).

Dataset for the study

For investigation of the AC-IE peer review process, information on a total of 1,899 manuscripts reviewed in the year 2000 was used. The information was drawn from material in the archives of the journal's publishing house, Wiley–VCH. In addition to internal review by the publisher, there were a total of 4,593 external reviews of the 1,899 manuscripts using an evaluation form and/or a comment sheet. For the statistical analysis in this study, the mean of the independent assessments of the importance of a submission was determined for each manuscript. According to Thorngate et al. (2009), the average error in ratings decreases with an increasing number of raters. Based on the external reviews, 46% ($n = 878$) of the 1,899 manuscripts were accepted for publication in AC-IE, and 54% ($n = 1{,}021$) were rejected. Research in the literature databases Web of Science (Thomson Reuters, Philadelphia, PA, USA) and Chemical Abstracts (CA, Chemical Abstracts Services, Columbus, OH, USA) revealed that 959 (94%) of the 1,021 rejected manuscripts were later published in other journals in a more or less revised form (Bornmann and Daniel 2008a, b, 2009a; Bornmann et al. 2009). Single publication *h* index values could be calculated for 1,814 of these manuscripts (876 accepted and 938 rejected but published elsewhere manuscripts).

Procedure for calculation of the single publication h index values for the accepted and rejected but published elsewhere manuscripts

The single publication *h* index and the citation counts for the accepted and rejected but published elsewhere manuscripts were determined using the Science Citation Index (Thomson Reuters). The data were captured from the Web of Knowledge Web Services XML Gateway, a Simple Object Access Protocol (SOAP) interface of Thomson Reuters providing access to the Web of Science (http://search.cpan.org/~timbrody/SOAP-ISIWoK-1.05/). The following steps were coded in a computer program: (1) The bibliographic data (author(s), journal title/name, volume, starting page, and publication year) of each of the 1,814 manuscripts were searched to identify and select the corresponding source records of the Web of Science; (2) For each selected source record, all citing papers were retrieved, including the times-cited information for each citing paper. The time window for all citation data covers the period from the year of publication to the date of searching (March 2010); (3) The single publication *h* index for the citing papers was computed from the list sorted by times-cited.

Statistical procedure

To test whether differences in the average single publication *h* index for two independent samples (accepted or rejected but published elsewhere manuscripts) were statistically significant, the Wilcoxon-Mann–Whitney test was computed. As there were more than two independent samples in the comparison of the different reviewers' rating groups, the

Kruskal–Wallis test was used instead (Sheskin 2007). The correlations between (1) citation counts and (2) reviewers' median ratings, respectively, and the single publication *h* index were computed using Spearman's rank correlation coefficient; the associations between the editorial decision by the AC-IE editors and (1) the citation counts and (2) the single publication *h* index, respectively, were computed using the point biserial correlation coefficient (StataCorp. 2009).

Results

Table 1 shows the minimum, maximum, arithmetic mean, standard deviation, and median of the single publication *h* index for manuscripts accepted by AC-IE or rejected but published elsewhere. The table presents the values for all publications and the values for papers published in 2000 and 2001. The manuscripts rejected by AC-IE were published elsewhere between 2000 and 2006; most of them in 2000 or in 2001 (see Bornmann and Daniel 2008b). As in a similar way to citation counts the *h* index value for a publication depends, among other things, on the size of the time window for the citing publications (and their citations), with the breakdown for the two publication years a standardization of the window is carried out.

As the results in Table 1 show, the *h* index values for the individual publications vary from 0 to 76. If a publication has an *h* index of 0, it is either uncited or it has citing papers that themselves have not been cited. If a publication has an *h* index of 76, it has been cited by at least 76 publications that themselves have been cited at least 76 times. Both the mean values (arithmetic average and median) for all publications and the mean values for only the publications in 2000 and 2001 show that manuscripts accepted by and published in AC-IE have higher *h* index values than publications that were rejected by AC-IE but published elsewhere. Whereas, for example, manuscripts accepted by and published in AC-IE in 2000 have a median *h* index of 14, the median *h* index of the manuscripts rejected by AC-IE and published elsewhere in 2000 is 11. The difference between the *h* index

Table 1 Minimum (min), maximum (max), arithmetic mean (mean), standard deviation (sd), and median of the single publication *h* index for manuscripts accepted by AC-IE or rejected by AC-IE and published elsewhere, broken down for year of publication

Editorial decision	*n*	Min	Max	Mean	sd	Median
All publications						
Accepted	876	1	76	16.79	10.69	14[§]
Rejected	938	0	56	11.70	8.40	10[§]
Total	1,814	0	76	14.16	9.90	12
Published in 2000						
Accepted	526	1	71	16.91	11.08	14*
Rejected	249	1	56	13.32	9.15	11*
Total	775	1	71	15.75	10.63	13
Published in 2001						
Accepted	350	1	76	16.61	10.07	14[$]
Rejected	531	1	45	11.70	7.99	10[$]
Total	881	1	76	13.65	9.19	12

Notes : [§] $z = -11.65$, $P < 0.05$, * $z = -4.72$, $P < 0.05$, [$] $z = -7.86$, $P < 0.05$

values for the two groups with different editorial decisions in Table 1 is statistically significant for all publications and also for publications in 2000 and 2001.

Table 2 shows the results on the association between the *h* index values and the median ratings by the reviewers. The reviewers ratings' range from 1 = very important to 4 = unimportant. As the mean and median values in the table show, a better rating by reviewers is usually associated with a higher single publication *h* index. Whereas, for example, publications with a median reviewers' rating of 4 show a median single publication *h* index value of 7.5, the median *h* index value of publications with a median reviewers' rating of 1 is about twice as high, at 14. Similar to the case with the editorial decisions, here again the differences between the median single publication *h* index values for the publications with different median ratings by the peer reviewers are statistically significant (for all publications and for papers published in 2000 and 2001).

Table 2 Minimum (min), maximum (max), arithmetic mean (mean), standard deviation (sd), and median of the single publication *h* index for manuscripts with different median reviewers' ratings on the importance of the manuscripts, broken down for year of publication

Median rating	*n*	Min	Max	Mean	sd	Median
All publications						
1	78	3	47	17.23	9.68	14.00*
1.5	117	2	71	16.59	11.39	14.00*
2	840	0	76	15.44	10.44	13.00*
2.5	266	1	54	13.18	9.32	11.00*
3	407	1	50	11.53	8.06	10.00*
3.5 or 4	50	1	43	10.90	9.37	7.50*
Total	1,758	0	76	14.22	9.95	12.00
Published in 2000						
1	38	7	40	16.87	8.05	13.50$
1.5	73	2	71	17.45	12.94	14.00$
2	415	1	71	16.63	11.04	14.00$
2.5	93	1	46	14.43	9.37	12.00$
3	131	1	50	13.21	8.85	11.00$
3.5 or 4	7	2	43	10.86	14.55	5.00$
Total	757	1	71	15.81	10.68	13.00
Published in 2001						
1	37	3	47	18.08	11.36	15.00§
1.5	42	3	42	15.40	8.20	13.00§
2	377	1	76	14.99	9.78	13.00§
2.5	145	1	42	13.00	8.85	11.00§
3	219	1	35	11.22	7.40	10.00§
3.5 or 4	28	1	35	11.25	9.15	10.00§
Total	848	1	76	13.71	9.21	12.00

Notes: Answer categories for the question about the importance of a manuscript: (1) very important; (2) important; (3) less important; (4) unimportant. Because of the small number of cases, the median reviewers' ratings of 3.5 and 4 were combined for the statistical analysis

* $\chi^2_5 = 76.1$, $P < 0.05$

$ $\chi^2_5 = 20.9$, $P < 0.05$

§ $\chi^2_5 = 35.7$, $P < 0.05$

The results reported above on the relationship between ratings by peer reviewers and the single publication *h* index values are in line with Bornmann and Daniel's (2008a, b) findings on the basis of citation counts for individual manuscripts that were accepted or rejected and published elsewhere. In Bornmann and Daniel (2008a, b) the differences in the impact values of the differently rated manuscripts were also statistically significant. The agreement of the results could be an indication that citation counts and the single publication *h* index are highly correlated. However, there are differences in this relationship depending on citation impact: For example, between times cited with fewer than 150 citations and the single publication *h* index the correlation coefficient is 0.95 ($n = 1{,}735$); between times cited with more than 149 citations and the single publication *h* index the coefficient is 0.53 ($n = 79$). The difference in the correlation coefficients could in fact mean, as Schubert (2009) maintained, that the additional consideration of indirect citation influence in the computation of the single publication *h* index for high impact publications contributes to a more refined picture of the performance of a single publication than with citation counts only. (Another explanation for the difference in the correlation coefficients could be that the higher the h index the more citations are needed to increase it, see here Franceschini and Maisano 2010). We want to test the extent to which this assumption is correct based on the editorial decisions and the reviewers' ratings of the individual publications.

To do so, first we assigned all publications ($n = 1{,}832$) to five percentile rank classes based on their citation counts (see here Bornmann et al. 2007; 2011). For the further analysis we selected the manuscripts ($n = 388$) in the class with the fewest citations (from 0 to 12 citations, median citation counts = 7) and the manuscripts ($n = 365$) in the class with the most citations (from 60 to 501 citations, median citation counts = 97) and determined the correlation between editorial decision or median reviewers' ratings and citation counts or single publication *h* index values. Just as with the previous evaluations, we again performed the analyses both for all publications and for publications from 2000 and 2001. Table 3 shows the results of the correlations with the editorial decisions, and Table 4 shows the results of the correlations with the reviewers' ratings.

Table 3 Coefficients (*r*) for the correlation between editorial decision (0 = rejection and published elsewhere; 1 = acceptance) and (1) single publication *h* index or (2) times cited for each publication, broken down for year of publication. The table shows the results for publications within all percentile rank classes, for publications with 0–12 citations (bottom 20th percentile) and for publications with 60–501 citations (top 20th percentile)

Correlation between editorial decision and …	All publications		Published in 2000		Published in 2001	
	r	*n*	*r*	*n*	*r*	*n*
Publications within all percentiles						
Single publication *h* index	0.26	1,814	0.16	775	0.26	881
Times cited	0.20	1,832	0.17	780	0.20	886
Bottom 20th percentile						
Single publication *h* index	0.14	370	0.13	157	0.11	191
Times cited	0.16	388	0.15	162	0.12	196
Top 20th percentile						
Single publication *h* index	0.16	365	0.06	150	0.19	177
Times cited	0.08	365	0.13	150	0.09	177

Table 4 Coefficients (r) for the correlation between median reviewers' ratings (here, the ratings were recoded from 4 = excellent to 1 = poor to avoid negative coefficients) and (1) single publication h index values or (2) times cited for each publication, broken down for publication year. The table shows the results for publications within all percentile rank classes, for publications with 0–12 citations (bottom 20th percentile) and for publications with 60–501 citations (top 20th percentile)

Correlation between reviewers' ratings and …	All publications		Published in 2000		Published in 2001	
	r	n	r	n	r	n
Publications within all percentiles						
Single publication h index	0.21	1,758	0.15	757	0.20	848
Times cited	0.20	1,774	0.17	762	0.20	852
Bottom 20th percentile						
Single publication h index	0.07	355	0.06	151	0.16	183
Times cited	0.12	371	0.12	156	0.20	187
Top 20th percentile						
Single publication h index	0.10	358	0.01	147	0.13	173
Times cited	0.09	358	0.09	147	0.05	173

As Table 3 shows, the results show a tendency in the right direction but are not consistent. With regard to the bottom 20th percentile (0–12 citations) there is a somewhat higher correlation between the times cited values and the editorial decisions than between the single publication h index values and the editorial decisions. Hence, with respect to the editorial decision, times cited values lead to a more refined picture than the single publication h index values. For the highly cited publications, the result is the reverse, as expected, for all publications and for the publications from 2001 (with a relatively clear difference between the correlation coefficients), but for the publications from 2000, contrary to expectations, the correlation between the times cited values and the editorial decisions is higher than the correlation between the single publication h index values and the editorial decisions. Table 4 shows the results for the median reviewers' ratings. Even though the size of the correlation coefficients is somewhat different here than the size of the correlation coefficients in Table 3, the results still show the same picture for the differences between highly cited publications and publications with few citations and between the single publication h index values and the times cited values. The results show a tendency in the right direction but are not consistent.

Discussion

According to Schubert (2009),

> citations place the cited paper into a network context, where its weight is determined not only by its node degree (number of citations) but also by some measure of its centrality (which depends on the weight of its citing papers, the weight of their citing papers, and so on). There are several possibilities to measure this; the h-index is one among them. (p. 564)

In this study, on the basis of assessments by peer reviewers we tested whether the single publication h index proposed by Schubert (2009) is in fact a useful yardstick to compare

the weight of different publications. As the results show (1) editorial decisions and (2) peer assessments are correlated with *h* index values: Manuscripts accepted for publication by the editors and with positive ratings by the reviewers show higher *h* index values after publication than manuscripts with negative ratings by reviewers and rejected by the editors but published elsewhere do. As this relationship had already been found when citation counts were used as the impact measure (Bornmann and Daniel 2008a, b), and as citation counts and single publication *h* index values are very highly correlated, as we found in this study, we tested whether the *h* index values for high impact publications yielded a more refined picture of the performance of a single publication than citation counts only (see here Schubert 2009).

Whereas the results of this analysis show a tendency in the assumed direction, they are not completely consistent. For this reason, we think that especially for testing the assumption it is necessary that further studies on the usefulness of the single publication *h* index be conducted for the group of the most cited papers. Does the new index proposed by Schubert (2009) in fact produce a more refined picture precisely in this impact area that is so important in evaluative bibliometrics? Only future studies with consistent results will determine whether Schubert's (2009) new index is an incremental contribution to the bibliometric toolbox in the high impact area or not.

Acknowledgments The entire research study, which is also investigating quality assurance of open access journals, is supported by a grant from the Max Planck Society. The authors would like to thank Dr. Christophe Weymuth (formerly at the Organic Chemistry Institute of the University of Zurich and now at BIOSYNTH AG, Switzerland) for investigation of the manuscripts rejected by *Angewandte Chemie International Edition* and published elsewhere.We thank Dr. Peter Gölitz, Editor-in-Chief of *Angewandte Chemie*, the Editorial Board of *Angewandte Chemie*, and the German Chemical Society (GDCh, Frankfurt am Main, Germany) for permission to conduct the evaluation of the selection process of the journal, and are grateful to the members of the editorial office for their generous support during the carrying out of the study.

References

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

Bornmann, L., & Daniel, H.-D. (2007). What do we know about the *h* index? *Journal of the American Society for Information Science and Technology, 58*(9), 1381–1385. doi:10.1002/asi.20609.

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. doi:10.1002/anie.200800513.

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. doi:10.1002/asi.20901.

Bornmann, L., & Daniel, H.-D. (2009a). Extent of type I and type II errors in editorial decisions: A case study on *Angewandte Chemie International Edition*. *Journal of Informetrics, 3*(4), 348–352.

Bornmann, L., & Daniel, H.-D. (2009b). The luck of the referee draw: The effect of exchanging reviews. *Learned Publishing, 22*(2), 117–125. doi:10.1087/2009207.

Bornmann, L., & Daniel, H.-D. (2009c). The state of *h* index research. Is the *h* index the ideal way to measure research performance? *EMBO Reports, 10*(1), 2–6. doi:10.1038/embor.2008.233.

Bornmann, L., & Daniel, H.-D. (2010). The manuscript reviewing process: Empirical research on review requests, review sequences and decision rules in peer review. *Library and Information Science Research, 32*(1), 5–12. doi:10.1016/j.lisr.2009.07.010.

Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., & Daniel, H. D. (2009). Convergent validity of bibliometric Google Scholar data in the field of chemistry. Citation counts for papers that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using

Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. *Journal of Informetrics, 3*(1), 27–35. doi:10.1016/j.joi.2008.11.001.

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. (2011). A multilevel modelling approach to investigating the predictive validity of editorial decisions: Do the editors of a high-profile journal select manuscripts that are highly cited after publication? *Journal of the Royal Statistical Society–Series A (Statistics in Society)*, *174*(4). doi:10.1111/j.1467-985X.2011.00689.x.

Bornmann, L., Weymuth, C., & Daniel, H.-D. (2010). 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.

Cole, S. (1989). Citations and the evaluation of individual scientists. *Trends in Biochemical Sciences, 14*(1), 9–13.

Franceschini, F., & Maisano, D. A. (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research, 203*(2), 494–504. doi:10.1016/j.ejor.2009.08.001.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569–16572. doi:10.1073/pnas.0507655102.

Schubert, A. (2009). Using the h-index for assessing single publications. *Scientometrics, 78*(3), 559–565.

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

StataCorp. (2009). *Stata Statistical Software: Release 11*. College Station: Stata Corporation.

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

van Eck, N. J., & Waltman, L. (2008). Generalizing the *h*- and *g*-indices. *Journal of Informetrics, 2*(4), 263–271.