

CITATION RANKING VERSUS EXPERT JUDGMENT IN EVALUATING COMMUNICATION SCHOLARS: EFFECTS OF RESEARCH SPECIALTY SIZE AND INDIVIDUAL PROMINENCE

C. Y.K. SO

*Department of Journalism & Communication, Chinese University of Hong Kong, Shatin, N.T. (Hong Kong)
E-mail: clementso@cuhk.edu.hk*

(Received August 25, 1997)

Numerous attempts have been made to validate the use of citation as an evaluation method by comparing it with peer review. Unlike past studies using journals, research articles or universities as the subject matter, the present study extends the comparison to the ranking of individual scholars. Results show that citation ranking and expert judgment of communication scholars are highly correlated. The citation method and the expert judgment method are found to work better in smaller research areas and yield more valid evaluation results for more prominent scholars.

Introduction

There has been some discussion about the validity of citation as a method for evaluating research significance. Numerous attempts have been made to validate this method as well as to compare it with other methods, notably peer review. Whereas past studies concentrate mainly on whether or not citation and expert judgment methods would behave differently in the evaluation of journals, research publications and universities, the present study extends the comparison to the evaluation of scholars. This study also makes further improvements on the comparison by providing richer details. In addition to an overall comparison, it also contrasts the two methods in terms of different sizes of research areas and different types of scholars within the same research area.

Specifically, the objectives of the present study are to (1) compare the citation method with the peer review method in the evaluation of individual scholars in the field of communication; (2) find out whether these evaluation methods work better in larger or smaller research specialty areas; and (3) see whether they perform better in evaluating more prominent scholars or less prominent ones.

Literature review

Comparison between citation measures and peer evaluation of performance was usually made using research institutions, research publications or journals as data. For example, a study of the quality of American universities found that peer rating review and bibliometric measures were highly correlated.¹ Depending on the specific fields concerned, the correlations between Roose-Andersen rankings and publication rankings were in the range of 0.65 to 0.91. However, this study found that university ranks and scores in different fields were much more highly correlated when based on peer assessment rather than on bibliometric measures. This means that when rating individual departments of a university people tend to be affected by the overall impression of the university. Bibliometric methods for rating individual departments may be more precise and free of the influence of this "halo effect." In the area of applied pharmaceutical research, *Koenig*² also found very high correlations between expert judgments and measures of publication activity among pharmaceutical companies. He also discovered that the bibliometric indicators performed slightly better than expert judgments in predicting research performance.³ For instance, the bibliometric measures could very well predict expert judgment variables but the reverse was not the case.

For research publications, *Lawani and Bayer*⁴ found in a study of 870 cancer research papers that even after controlling for factors like self-citations, language and country of authorship, highly rated papers were more highly cited than average papers. Both peer assessments and citation rates are said to be "reasonably reliable and largely interchangeable indices for evaluations in the aggregate," but they may not be able to differentiate between contributions from individual scholars.⁴

Another interesting comparison is between bibliometric measures and scholars' own judgments of their publications. *Porter, Chubin and Jin*⁵ found that in 107 papers by 28 award-winning chemists, 52 were listed as the most-cited papers, 30 were rated by the chemists themselves as their individual best papers, and only 25 were classified as both most-cited and best papers. There is a correlation between best papers and most-cited papers but the relationship is not perfect.

For evaluation of journals by the bibliometric method and peer assessment method, the two processes were found to be highly correlated in biomedical fields⁶ and social sciences⁷ but less so in the physical sciences.⁸

Other than using institutions, research articles and journals as the test subjects, some studies compared bibliometric data and survey data in terms of other characteristics of scientific research. For example, *Garrison, Merman and Lipton*⁹ found that

characteristics of subject matter and researchers' institutional sectors were similar in both types of data but funding sources were somewhat different.

As stated above, past comparisons between citation method and peer judgment method seldom focus on individual scholars. One reason is that when an individual is to be evaluated for faculty hiring, tenure or promotion, only that person's citation records and others' evaluation of him/her are scrutinized. It is difficult to systematically assemble many such individual cases for a more representative comparison. In contrast, the data for comparing journals, research articles or institutions as subjects are easier to retrieve. The present study hopes to fill this gap by comparing individual scholars' rankings in citation performance and peer evaluation.

Furthermore, past comparisons often focus on different academic fields and only report whether the correlation is high in some fields and not in others. Little discussion of the results is offered in terms of the different characteristics (such as size and nature) of these fields. In this paper the size of various research areas within the field of communication is studied to see whether it affects the performance of the two evaluation methods.

Data source

There are two sources of data in this study--citation and expert survey. The citation data are from the 1985, 1986 and 1987 annual conferences of the International Communication Association (ICA). A total of 1,319 conference papers representing 13 divisions and special interest groups of communication (see Table 1 for details) were collected from these three conferences. There are all together 44,877 citations found in these papers. Although the ICA conference papers may vary in quality, as one big data set it is highly representative of the works done in the field. Many important research areas in the field of communication are covered by the ICA conference and thus a comparison of the citation characteristics among these areas can easily be made.

The objective of gathering the bibliometric data from these ICA papers is to identify the most cited authors in communication studies. From the citation data a list of most cited authors can be constructed for each of the 13 ICA divisions and special interest groups. The length of the lists for different research areas in communication differs since cutting points for inclusion are different. In general, those ICA divisions with longer history and more membership (such as mass communication and interpersonal communication) have a long list of most cited authors while the relatively new ones (feminist scholarship, popular communication) have a short list. Then, from each of the 13 lists 10 scholars were selected by means of systematic random sampling. These 10

scholars, with varying amounts of citations received, can represent the whole spectrum of citation prominence in that division/special interest group. Taking a constant-sized sample from ordered populations of widely varying sizes can create problems. For example, this approach may have forced a mechanical comparison among scholars from different divisions. However, it seems that the present approach enables the comparison to be made in a relatively simple and direct way without serious loss of validity.

The second data set is from a questionnaire survey of a selected group of scholars in communication. In order to evaluate the quality of communication research studies, expert opinions should be sought instead of the views of the general members of the academic community. This study taps the opinions from two groups of people who are more knowledgeable about the field of communication. The first group is composed of most cited authors (determined by the first data source) in various areas of communication studies, and members of the second group are key administrators in numerous academic institutions in communication (including all office-bearers of ICA, SCA (Speech Communication Association) and AEJMC (Association for Education in Journalism and Mass Communication), communication department/school heads from 26 major universities, and 31 journal and serial publication editors). A total of 216 most cited authors and 238 key administrators were selected. Discounting the overlap between these two groups of people, 399 people were sampled. A questionnaire was mailed to them in early 1991 and 234 usable questionnaires were returned within two months, representing a response rate of 58.6%.¹⁰

In the questionnaire, respondents of a certain division/special interest group were asked to rank the 10 selected scholars in that particular division/special interest group. These expert judgment rankings were compared with the citation rankings obtained from the first data source. If the rankings generated by the citation data correlate highly with the rankings of the expert survey data, the validity of citation method can thus be established.

Results and discussion

Table 1 shows the various results of comparison between the two methods. In the Table, the average difference of rank refers to the results obtained by the citation method and peer review method among the 10 selected scholars in each division/special interest group. A smaller figure in the average difference of rank means that the two evaluation methods tend to agree with each other in that particular division/special interest group. For example, citation and expert rankings in MC and ISD show larger

discrepancies while in PR and PH the discrepancies are smaller. A difference of the means test for independent samples is performed in each division, with the results shown in the T value and T significance columns of Table 1.

The next three columns are indicators of the relative size of the 13 divisions/special interest groups. First, the "N of citation list" column reveals the position of the 10th author which in effect signifies the sample size from the citation data in each division/special interest group. The "N of survey" column indicates the sample size of the expert questionnaire survey in each division/special interest group. They should to a certain extent correspond with the size of ICA membership. For example, MC, IP and OC are the three largest divisions and PO, PR and FS are the three smallest.

Table 1
Comparison of citation evaluation and expert evaluation on highly cited authors
in the field of communication

Division*	Average difference of rank**	T value & significance	N of citation list	N of survey	ICA membership
IS	1.0	10 (n.s.)	73	29	653
IP	2.0	25 (n.s.)	85	46	1019
MC	2.8	25 (n.s.)	82	64	1480
OC	2.0	26 (n.s.)	87	22	1169
ICD	1.4	18 (n.s.)	74	31	972
PC	1.8	25 (n.s.)	73	23	682
ISD	2.6	24 (n.s.)	51	21	506
HC	1.4	22.5 (n.s.)	45	12	564
PH	0.8	23 (n.s.)	28	10	711
CAT	2.2	24 (n.s.)	58	19	581
PO	1.2	22.5 (n.s.)	28	6	398
PR	0.6	17 (n.s.)	28	15	423
FS	1.7	24.5 (n.s.)	29	9	500

*Names of the divisions/special interest groups in the field of communication are as follows: IS = Information Systems, IP = Interpersonal Communication, MC = Mass Communication, OC = Organizational Communication, ICD = Intercultural and Development Communication, PC = Political Communication, ISD = Instructional and Developmental Communication, HC = Health Communication, PH = Philosophy of Communication, CAT = Communication and Technology, PO = Popular Communication, PR = Public Relations, FS = Feminist Scholarship.

**The average difference of rank for a particular division/special interest group is calculated by averaging the differences of ranks for the 10 individual most cited authors. For each author, the difference is calculated by comparing his/her ranks in the citation list and the survey result list. The differences here are expressed in terms of absolute numbers.

In Table 1, the average differences of rank for all the divisions/special interest groups are not significant. This suggests that the results generated by the citation method and the peer review method are basically the same. However, a closer look at the results of each of the 13 divisions/special interest groups reveals that the citation method is actually better in terms of identifying and ranking the scholars. While most of the scholars in a division obtain a similar ranking from the two methods, a few of them are quite "misplaced" in the peer review method. The lack of face validity as judged from careful scrutiny may be the outcome of the difference in the actions taken. The act of citing is usually better than mere naming of people, as making a citation is done in the context of a more "serious" task of writing a formal academic paper. Peer review through questionnaire, on the other hand, often involves a more informal manner. As the attitude in peer reviewing is comparatively more "loose," a less well-known person in a field might be improperly ranked or even mistakenly identified with another person who shares a similar last name. In the division of MC, for example, Horace Newcomb was ranked ninth in the citation method but third in the questionnaire evaluation. This substantial discrepancy might be due to the fact that some questionnaire respondents mistook him for Theodore Newcomb, who is probably a better-known figure in mass communication. In some cases a well-known scholar in another field might be wrongly perceived by questionnaire respondents as an important figure in communication simply because that scholar is so famous in the social sciences. The case of Piaget is a good illustration. In the division of ISD, Piaget was ranked only fifth, which more or less faithfully reflects his importance in this area. But he was ranked first in the questionnaire results for this division.

In reviewing the comparison between the citation method and the peer judgment method, a conceptual issue worthy of attention is to what extent are they measuring the same thing. Citation counts may emphasize more on recent publications while peer judgment tends to take a person's whole academic career into consideration. Thus, the length of time scholars working in the field can be a crucial factor. This difference may explain the case of Piaget. Hence, further analysis is needed to probe into the nature and implications of the two methods.

Similar results from the citation method and the peer review method for the 13 divisions and interest groups show that they work reasonably well in evaluating scholars' importance regardless of the size of the academic community. However, in relation to the second objective of this paper, it would be desirable to know which method performs better in different contexts of size. In Table 1, the first six divisions (IS through PC) are larger while the others are smaller in size. A little calculation shows that the mean of average difference of rank for the first six divisions is 1.83 and that for

the other (smaller) divisions it is 1.50. If we compare the three largest divisions (MC, IP and OC) and the three smallest (PO, PR and FS), the difference (2.27 versus 1.17) is even more pronounced. As stated above, a smaller average difference implies steadier performance between the two methods. Thus, citation evaluation and peer review may work better in the context of smaller divisions which yield more stable results. This indirectly lends favorable support to the practice of evaluating research specialties (not the whole field), since natural and social sciences are structurally organized in terms of numerous small research specialties.¹¹

Table 2
Average difference of rank by author position

Division	Overall average (10 authors)	Average of 1st-3rd authors	Average of 4th-7th authors	Average of 8th-10th authors	Trend
IS	1.0	1.0	1.8	0.0	
IP	2.0	1.0	2.3	2.7	up
MC	2.8	1.7	3.0	3.7	up
OC	2.0	1.0	3.0	1.7	
ICD	1.4	1.0	1.3	2.0	up
PC	1.8	1.7	2.8	0.7	
ISD	2.6	3.3	2.5	2.0	down
HC	1.4	0.3	1.8	2.0	up
PH	0.8	1.0	0.3	1.3	
CAT	2.2	2.3	1.3	3.3	
PO	1.2	0.7	1.8	1.0	
PR	0.6	0.7	1.0	0.0	
FS	1.7	0.8	1.8	2.5	up
Average	1.65	1.27	1.87	1.76	

The third objective of this study is to find out whether the two methods differ in the evaluation of scholars with different degrees of prominence. Here the 10 most cited authors in each division/special interest group are broken down into three groups, with the first three scholars in the first group, the fourth to seventh scholars in the second group, and the last three scholars in the third. A comparison can then be made of the average difference of rank among these three groups. In Table 2, the first column is the overall average difference of rank (with 10 authors in each division/special interest group). The subsequent columns are average difference of rank for the three groups. The last column indicates whether there is an upward or downward trend with respect to the size of the average difference of rank when moving from the more cited authors to the less cited ones.

At the bottom of Table 2, the respective average figures of average difference of rank for the three sets of scholars are 1.27, 1.87 and 1.76. The first three scholars clearly show a smaller difference in the ranks obtained from the two evaluation methods. In other words, the top scholars in various divisions/special interest groups generally have a lesser chance of being mis-identified or mis-ranked. Conversely, caution must be duly exercised in the evaluation of less prominent scholars by means of citation or peer review.

The same conclusion can be reached by examining the trends in individual divisions/special interest groups. Among the 13 research areas in communication, 5 show an upward trend (increasing difference of rank between the two methods), 1 shows a downward trend, and the rest show no trend. These figures point out that as we move from more prominent scholars to the lesser ones, discrepancies between the two methods gradually widen.

To conclude, there is evidence that citation data can be applied to the evaluation of universities, journals, research publications and also individual scholars. The validity of the citation method is once again demonstrated. As citation results correlate highly with expert review results, we are more comfortable in the use of citation data as an alternative and even a substitute for peer review exercise. In fact, citation data are comparatively easier to obtain, quantifiable and more objective in nature.¹²

Results of this study as well as others¹⁻³ even suggest that in some aspects citation data perform better than subjective peer evaluations. Although more studies are needed before we can arrive at such a conclusion, at present we should at least feel more confident about the validity of the citation method, especially when we are evaluating more prominent scholars in the context of smaller research specialty areas. For many practical purposes such as faculty hiring and promotion, only a few candidates are involved whose interests fall in specific research areas. Citation measures can provide valid, objective and comparative indicators for evaluation. The citation method is becoming a more popular evaluation tool in addition to traditional peer review.

Notes and references

1. R. C. ANDERSON, F. NARIN, P. MCALLISTER, Publication ratings versus peer ratings of universities, *Journal of the American Society for Information Science*, 29 (1978) 91-103.
2. M. E. D. KOENIG, Determinants of expert judgement of research performance, *Scientometrics*, 4 (1982) 361-378.
3. M. E. D. KOENIG, Bibliometric indicators versus expert opinion in assessing research performance, *Journal of the American Society for Information Science*, 34 (1983) 136-145.

4. S. M. LAWANI, A. E. BAYER, Validity of citation criteria for assessing the influence of scientific publications: New evidence with peer assessment, *Journal of the American Society for Information Science*, 34 (1983) 59-66.
5. A. L. PORTER, D. E. CHUBIN, X. JIN, Citations and scientific progress: Comparing bibliometric measures with scientific judgments, *Scientometrics*, 13 (1988) 103-124.
6. P. R. MCALLISTER, R. C. ANDERSON, F. NARIN, Comparison of peer and citation assessment of the influence of scientific journals, *Journal of the American Society for Information Science*, 31 (1980) 147-152.
7. M. D. GORDON, Citation ranking versus subjective evaluation in the determination of journal hierarchies in the social sciences, *Journal of the American Society for Information Science*, 33 (1982) 55-57.
8. A. SINGLETON, Journal ranking and selection: A review in physics, *Journal of Documentation*, 32 (1976) 258-289.
9. H. H. GARRISON, S. S. HERMAN, J. A. LIPTON, Measuring characteristics of scientific research: A comparison of bibliographic and survey data, *Scientometrics*, 24 (1992) 359-370.
10. The citation data collected in 1985-87 and the questionnaire data collected in 1991 were both for the author's dissertation research. Collecting both types of data simultaneously should be more ideal, but the time lapse between the two sets of data should not pose a major threat to validity because overall citation patterns as well as peer judgment usually remain quite stable over a short period of time. For more details on the data collection methods, see C. Y. K. SO, *Mapping the Intellectual Landscape of Communication Studies: An Evaluation of Its Disciplinary Status*, Unpublished Ph.D. dissertation, Annenberg School for Communication, University of Pennsylvania, 1995 (UMI Number 9615129).
11. D. CRANE, Social structure in a group of scientists: A test of the "Invisible College" hypothesis, *American Sociological Review*, 34 (1969), 335-352.
12. E. GARFIELD, *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*, ISI Press, Philadelphia, 1979.