Information Processing & Management Vol. 25, No. 6, pp. 713–719, 1989
Printed in Great Britain.
0306-4573/89 $3.00 + .00

VISIBILITY OF PERIPHERAL JOURNALS THROUGH THE SCIENCE CITATION INDEX*

LJERKA ZMAIĆ and SINIŠA MARIČIĆ
The Research Library of the National and University Library, Zagreb, Croatia, Yugoslavia

and

VLADIMIR SIMEON
Laboratory of Physical Chemistry, Faculty of Science, University of Zagreb, P.O.B. 163, 41000 Zagreb, Croatia, Yugoslavia**

(*Received* 21 *October* 1988; *accepted in final form* 23 *January* 1989)

Abstract—The effect of the inclusion of a journal into the SCI source journals selection on the journal's "visibility" was studied by analyzing the number of independent citations of the articles published in two journals (*Croatica Chemica Acta* (CCA) and *Roczniki Chemii* (RC)) in two periods. The "echo factor," i.e. the citation count normalized by the size of CCA and RC publication source pools and of SCI citing pool, did not show any increase upon the inclusion of either of the two journals into SCI selection: for CCA it even showed a decrease and for RC remained essentially constant. A statistically significant increase in the SCI citing pool, that took place around 1976, was also noticed.

INTRODUCTION

The most important role of journals published by (geopolitically) small scientific communities is to establish communication ties with the world's scientific endeavor. For the learned societies of such peripheral science (for the meaning of this term see [1] and [2]) to foster this role of their journals means getting them into the mainstream of scientific information. Before the appearance of the Science Citation Index (SCI) [3] the abstracting services were almost the only channels through which such a role of "small" journals could be accomplished. The additional advantage of SCI is two-fold: (i) it is searchable from an interdisciplinary point of view [4,5], and (ii) it records, through the citations, even those papers that may have been published in journals not included in the SCI selection for regular processing [6].

However, it is a common-sense reasoning that more citations will be recorded of articles from a journal once it is selected for regular processing in SCI. This notion is so widespread within the peripheral science that the editors have a hard time squeezing their journals into SCI.

It therefore seemed worthwhile to do a piece of research on whether a journal becomes more "visible" upon its acceptance for regular processing in SCI. With that formulation of the research task we are within the realm of scientific information field *per se*. The other side of the coin is that the results of this line of research may give additional insight into the soundness of citation analysis when used in science-of-science studies of the peripheral science [2,7].

METHOD

Choice of journals

From 1961 till 1963, i.e., within the first three years of the SCI's existence, no Yugoslav journals were included among some 600 journals in the SCI selection. In 1964, some

*Part of the M. Sc. Thesis by Lj. Z.
**Address for correspondence.

700 journals were being processed by SCI and *Croatica Chemica Acta* (CCA—formerly *Arhiv Za Kemiju*) was included among them.

In looking for a comparable journal from another country we opted for *Roczniki Chemii* (RC) from Poland, included in SCI since 1964, too. Both journals were reasonably "seasoned" by the time they were included in SCI: CCA had been published since 1927 (with its present title and editorial style since 1956), and RC since 1921. Both of them had been edited according to international standards.

Source volumes

The main difference in publishing terms between CCA and RC is the number of pages per each volume/year: the latter was about three times as voluminous as CCA. That is the reason we used three years before (1961-1963) and after (1972-1974) its inclusion in SCI (1964) for CCA, whereas only one year (1964 and 1975, respectively) was sufficient for RC. (For number of papers in each case see Table 1.)

Citation periods

The first period of citation we call the "dynamic" one, because the SCI database expanded within it from some 600 to 2500 journals. The citation data from that period, i.e. for papers published before the journal had been included into the SCI database, served for comparison with the citation rates in the second, "steady-state" period while SCI maintained an almost constant database (with a very slow increase) and the journal had already been included in SCI.

Citation types

We are not interested here in the disputes about the meaning of citations (the interested reader may consult yet another contribution to this "field" [2]). The reason for such a stand is that, whatever the origins and the consequences of citation practices might be, we are interested here only in the visibility of (peripheral) journals through the SCI as a communication means of the contemporary science information system.

However, it is desirable to exclude at least one type of citations that cannot be regarded to be due to SCI itself, and that is the so-called self-citation. What is recorded and discussed here are independent citations.

The latter comprise all citations from any journal (in the SCI database) except: (a) citations from either CCA or RC, and (b) citations from the papers having one or more authors in common with the source article. The latter restriction was possible to observe more stringently in the case of CCA than for RC.

Namely, the "self-citations" may be defined in several ways [8]. In this study we discriminated against the following types of self-citations:

(i) citations by first authors who were also authors or co-authors of the source article
(ii) "hidden" self-citations that were established upon inspection of the corresponding source index of SCI, namely whether any of the co-authors of the citing article (i.e., not only the first one) was a co-author of the source (cited) paper
(iii) citations by authors publishing papers with the members of the research group with which the author of the cited paper publishes, too, although none of the authors or co-authors of the cited article and the citing one are identical
(iv) citations from papers published by the source journals (CCA and RC)

Although all the four kinds of self-citations were excluded for CCA, it has not been possible to exclude type iii for RC. However, bearing in mind that we compare two periods of citations collected in exactly the same manner (though differently for CCA and RC), this point should have little or no influence in drawing the final conclusions.

An overview of the number of source papers, the altogether cited papers and the number of independently cited ones is given in Table 1.

Table 1. The basic data: number of papers published, cited (in total) and cited only independently

	Publishing year			
	CCA		RC	
	1961–1963	1972–1974	1964	1975
Papers published	117	169	250	267
Cited papers (in total)	92	136	162	217
Independently cited papers	61	96	103	159
Period of citation	1961–1969	1972–1980	1964–1974	1975–1984

Errors due to SCI [4]

The accidental errors were quite numerous, originating in incorrect references given by the authors in the source articles processed by SCI, or being due to misprinting in SCI. They were mostly eliminated in the course of recording the data for this study, such as the names of the authors (when given only surnames and/or surnames plus one or more initials, or with incorrect initials). Out of 286 articles (for CCA) there were 21 articles requiring a later, more elaborate correcting procedure: most frequently the pagination was wrong (11 cases), volume or year (6), page, volume, and year (2), and changed order of the co-authors (2). The incorrect year could influence mostly this type of study—if the year recorded was outside (i.e. not within) the chosen citing period, the citation could not be taken into account. It is not known how many of such errors may have crept in altogether, but some of the citations of that type discovered later were incorporated in the overall data.

The systematic errors in SCI are two-fold: (a) the well-known problem of recording only the first author of the citing paper (we resorted, as already mentioned, to the source index in looking for self-citations), and (b) the choice of the journals for the SCI database. Although the latter point is of interest in other respects, especially concerning the peripheral science, it is not expected that any fluctuation in the SCI selection of journals could introduce serious errors into this type of study.

Let us point out that we are comparing the citations from two periods. Hence, whatever errors may have crept in it is reasonable to expect that substantial differences in the contributions of each type of error in such a comparison are most unlikely. The figures given for the accidental errors suggest that the uncertainty introduced by both kinds of errors is of the order of 10%. Neither can one expect that the citation practices could have drastically changed within the investigated period.

Normalisation of data

For lack of any accepted citing theory, and of convincing evidence to the contrary of the initial hypothesis, we made the following assumptions:

a. Within about 20 years encompassed by the present study, there was no substantial alteration of citation motives (especially with regard to peripheral journals).
b. The relationship between the number of source articles and the pool of citing journals is linear.

Hence, the following normalizing formula was constructed:

$$N_e = (N_{ind,a})_p / (C_p S_a) \quad (1)$$

where N_e is the "echo" factor, i.e. the normalized number of received independent citations (for a given year of publication and citation); $(N_{ind,a})_p$ is the number of independent citations received by source papers (published in the year a) in the year p; C_p is the size of the citing pool in the year p; and S_a is the size of the source pool in the year a.

Statistical evaluation of data

Nothing being known on the distributional properties of the analyzed quantities, non-parametric statistical methods (Mann-Whitney test, Kruskal-Wallis analysis of variance [9]) were used whenever possible. SPSS/PC program package was used for all statistical computations.

RESULTS

The citation raw data are summarized in Table 2. The yearly number of independent citations is given for both journals (CCA and RC), for the two periods of publishing the

Table 2. The independent citation counts and corresponding echo factors (in parts per million)

	Publishing year							
	CCA						RC	
Citing year	1961	1962	1963	1972	1973	1974	1964	1975
1961	1 0.25							
1962	1 0.23	0 0.00						
1963	6 1.32	6 1.19	2 0.36					
1964	17 3.20	25 4.22	11 1.68				10 0.26	
1965	13 1.57	16 1.73	11 1.08				12 0.20	
1966	7 0.73	6 0.56	16 1.35				28 0.40	
1967	11 1.03	10 0.84	8 0.61				23 0.30	
1968	4 0.37	1 0.08	14 1.05				23 0.29	
1969	7 0.58	14 1.05	16 1.08				29 0.33	
1970							20 0.22	
1971							38 0.41	
1972				3 0.12			20 0.21	
1973				11 0.37	0 0.00		27 0.26	
1974				23 0.92	6 0.20	0 0.00		
1975				21 0.80	30 0.99	0 0.00		0 0.00
1976				13 0.46	11 0.33	10 0.63		14 0.11
1977				9 0.29	15 0.42	5 0.28		37 0.28
1978				12 0.38	10 0.27	10 0.57		39 0.29
1979				9 0.28	19 0.50	11 0.60		39 0.28
1980				4 0.12	15 0.40	3 0.16		37 0.26
1981								31 0.21
1982								28 0.19
1983								33 0.21
1984								21 0.13

source articles (1961–1963 and 1972–1974 for CCA and 1964 and 1975 for RC) with the two periods of citation – dynamic and the steady-state (1961–1969 and 1972–1980 for CCA; 1964–1973 and 1975–1984 for RC). The values of echo factor (in parts per million) are given below the corresponding citation counts. Echo factors (N_e) were calculated from eqn (1) (see Discussion), using the following data: $(N_{ind,a})_p$ are the absolute citation counts (Table 2); C_p is the number of the articles in the SCI citing pool, according to SCI data [10]; S_a is the publication source pool (Table 1).

DISCUSSION

Citing pool

As already mentioned, the volume of SCI database [10] was steadily growing, the changes in the number of citing journals being especially noticeable in the first decade of the scanned period (1961–1971). The growth pattern of the number of articles contained in the citing journals was roughly similar, although the rates differed to some extent in the "dynamic" and the "steady-state" periods. The pattern of growth of the citations themselves, though similar by and large, exhibited a jump in the year 1976, thus raising on the whole the level of citation in subsequent years.

All this was examined in more detail by statistical methods in order to find an adequate normalising procedure (see Methods section). The data (YP = the year of publication, NJ = number of journals in SCI selection, NA = number of articles contained in them, NC = number of citations they received) were taken from ref. [10] for the 1961–1981 time interval.

Product-moment correlations between these four quantities were found to be high (>0.932) and the diagonalization of the correlation matrix revealed only one statistically significant principal component accounting for as much as 97.7% of total variance.

The significance of the abrupt change in citation rates in 1976 and their higher level thereafter was checked by regression analysis, using the dummy-variable technique: a binary variable was introduced which was assigned the value of 0 (zero) in the 1961–1975 period and the value of 1 in the subsequent period (1976–1981). Irrespective of the strategy of inclusion/exclusion of the variables treated as independent (YP, NJ, NA, dummy) only three of them were identified as significant at high confidence levels (student's test of regression slopes gave $P \leq 0.0231$ and $R^2 = 0.9897$), *viz.* time (YP), number of articles contained in the citing journals (NA) and the dummy variable reflecting the jump in the citation rate. (The number of journals (NJ) was not found significant in this respect.) The significance of the regression can be judged from its analysis of variance: the regression of citation rates on YP, NA, and dummy variable yielded an F ratio (explained/residual variance) of 639.3 (with 17 and 3 degrees-of-freedom) which corresponds to a confidence level of $P < 0.0001$. The standardized regression coefficients (beta weights), which are of less importance in the present context, ranged from 0.2 to 0.5. From the value of the squared correlation coefficient (R^2) it is seen that three predictors (YP, NA, dummy) accounted for as much as 99% of the observed values of citation rates, the residual variation being thus mere 1%. So precise a linear regression indicates that the observed jump in the citation rates was not a fortuitous outlier but rather a (roughly constant) feature of the year 1976 and subsequent years where the citation rates were consistently higher than in the preceding years. This jump deserves further study which is outside the scope of the present work. It may, namely, reflect the onset of distinctly different citation practices. This finding seems to deserve a more systematic investigation.

Data normalisation

It seems plausible that the yearly number of (independent) citations should be proportional to the number of the source articles being studied and, also, to the size of the pool of the citing journals (because of the high correlation between these two quantities).

Strictly speaking, the citing pool should have been defined as the number of potentially citing articles (to be determined a posteriori), i.e. as the number of all articles published by journals citing the one under study. Although this was done for CCA such data were not collected for RC. However, as the comparison of the citing pool growth curves did not show any substantial difference between the CCA citing journals increase and the total SCI pool, we resorted in what follows to the data given by SCI for the total number of articles in their journals per each year.

Our "echo" factor (N_e) is similar to the impact factor used by Garfield [11], but the two are not dimensionally identical. Garfield's impact factor is defined arbitrarily, i.e. it, too, is not founded theoretically. The impact factor is the ratio of the number of citations (received in a given year by papers published in the preceding two) and the number of source papers (published in the preceding two years). It does not take into account the growth rate of the citing pool (when a longer period is studied) and was therefore deemed unsuitable for the present study. The term "impact" is also value-laden, as it implies a definite consequence of the citation event although it is not known what the cause and effect here are. Our "echo" factor is neutral in that respect, implying simply an echo (to the published source).

Comparison of the two periods

For CCA, the data can be divided, according to the year of publication, into three groups for each of the two periods. By applying Kruskal–Wallis nonparametric analysis of variance (echo factor versus the year of publication) it was found that the three groups are homogeneous within each period ($P = 0.71$ and $P = 0.85$, respectively), but the data for the two periods are not ($P = 0.03$). The same conclusion was reached by applying two additional tests on the pooled data for each period: Mann-Whitney ($P = 0.0008$) and two-sample Kolmogorov–Smirnov ($P = 0.002$). Thus, the "visibility" of CCA, as measured by its echo factor, did in fact decrease in the later period, as compared to the earlier one.

Because of their simpler structure the data for RC were analyzed only by using Mann-Whitney and Kolmogorov-Smirnov tests, yielding the confidence levels of $P = 0.065$ and $P = 0.218$, respectively. Therefore, the null hypothesis (of no difference in echo factor between the two periods) was retained as valid. If the result of the Mann-Whitney test were treated as (marginally) significant then the echo factor for RC would also be lower after its inclusion into the SCI selection.

It may be noted in passing that the echo factor for RC, although on the whole much smaller than for CCA ($P = 0.00005$), remained much more uniform throughout the period investigated: Kruskal-Wallis analysis of variance (echo factor versus year of citing) gave a nonsignificant result ($P = 0.456$), in contrast to CCA where the echo factor varied significantly from year to year ($P = 0.022$). Could that be somehow related to the more stable science policy in Poland than in Yugoslavia?

These results show convincingly that the inclusion of *Croatica Chemica Acta* (CCA) and *Roczniki Chemii* (RC) into the Science Citation Index did not improve (i.e. did not increase) the visibility of these two journals within the world's scientific literature.

In case of CCA both the subject matters covered in the two periods and (consequently) the citing pool of journals were quite different. No such analysis was performed for RC, but with or without a similar change for RC, too, these results are very indicative of the fact that a (chemical) journal retains its identity for quite a long time, irrespective of whether it is or it is not in the SCI journal pool.

If further investigations with comparable journals but from other subject fields confirm, presently the only available results (for CCA and RC), it would mean that such journals (from peripheral scientific communities) do not depend on SCI for their inclusion into the world's scientific literature. This, of course, does not mean that SCI is superfluous for retrieving purposes, i.e. "to ease the opening of the already existing doors." However, an analysis [12] of the CCA impact factor in the later part of the steady-state period showed a considerable decline attributable to a corresponding decline in scientific quality (as evaluated independently). No doubt, it is only the scientific value of a "small" journal that can "keep it alive" within the world's science.

In addition to checking with journals from different subject fields, another point should also be examined, namely, there are "small journals" that were established much later than the two examined here, so that such journals may have not been followed up by secondary information services (other than SCI) long enough to bring them to the permanent attention of interested authors. It may turn out that for such younger journals from the scientific periphery inclusion into SCI database is important with respect to their "visibility."

A reply to the latter question is of importance in clearing the dilemma of whether it would be better if all the peripheral journals were left out of the SCI selection [13]. It would also add weight to the science-of-science studies of peripheral science through SCI records, because it might dispel the notion that SCI gives a skewed picture of the peripheral science owing to lack of the corresponding journals in their database selection. Until answers come along any analysis of peripheral scientific output using SCI should provide checks that the relevant peripheral source journals are not under the influence of the "SCI force-field" in attracting citations.

REFERENCES

1. Solte-Heiskanen, V. Evaluation of scientific performance on the periphery. Science and Public Policy, 13: 83-88; 1986.
2. Ferligoj, A.; Maričić, S.; Pifat, G.; Spaventi, J. Cluster analysis of citation histories from an institutional setting. In: Tudor-Šilovic, N.; Mihel, I., editors. Information research. Research methods in library and information science. London: Taylor Graham; 1988: 174-200.
3. Garfield, E. Citation indexing. New York: John Wiley & Sons; 1979.
4. Narin, F. Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity. Washington, D.C.: Computer Horizons, Inc.; 1976: 147.
5. Weinstock, M. Citation indexes. In: Encyclopedia of librarianship and information sciences, vol. 5, New York: Marcel Dekker; 1971: 16-40.
6. Spaventi, J.; Tudor-Šilović, N.; Maričić, S.; Labus, M. Bibliometrijska analiza znanstvenih časopisa iz Jugoslavije. Informatologia Yugoslavica, 11: 11-23; 1979.
7. Maričić, S. Evaluating the scientific merit of journals within the scientific periphery. In: Kuhnhold, W.W.; Kirchner, W.P., editors. Publications as an integral part of scientific research. Proceedings of the fifth international conference of scientific editors, Hamburg 1987. Hamburg: International Federation of Scientific Editors Associations; 1988: 61-74.
8. Tagliacozzo, R. Self-citations in scientific literature. Journal of Documentation, 33: 251-265; 1977.
9. Cooper, B.E. Statistics for experimentalists. Oxford: Pergamon Press; 1969: 273ff.
10. Science citation index, Annual guide and lists of source publications. Philadelphia: ISI, Institute for scientific information; 1987.
11. Garfield, E. Bradford's law and related statistical patterns. Current contents, No. 2, 1980: 5-12.
12. Zmaić, Lj. Croatica chemica acta i citation index. M. Sc. Thesis, The Centre for Postgraduate Studies, University of Zagreb, 1983.
13. Maričić, S. Scientific journals: Selection criteria for information services. Bulletin (45) Society for Social Studies of Science, 5: 6-8; 1980.