

New Tools for Improving and Evaluating The Effectiveness of Research

Irving H. Sher

and

Eugene Garfield

*Institute for Scientific Information
Philadelphia, Pennsylvania 19106*

ABSTRACT

Using a novel concept called citation indexing, the Institute for Scientific Information comprehensively indexes on a current basis every item published in over a thousand journals as well as every U.S. Patent issued. With this technique, the indexing terms are the references cited in the bibliographies of the current literature.

The *Science Citation Index* is published quarterly with annual cumulations and is used for retrospective searching. ASCA (Automatic Subject Citation Alert) is an analogous current alerting service whereby each user receives personal weekly computer listings mailed in answer to continuing questions entered as individual profiles of interest. Both the printed indexes and the computerized alerting service conveniently inform the user of those items from the current literature which cite, in their bibliographies, known published works of interest to the user. The conceptual link established to earlier related work by a current author using bibliographic citations becomes the thread of continuity that can lead the user forward in time from any reference question, phrased simply as a citation, up to the pertinent current literature of science and technology. The unique citation indexing approach bypasses the usual semantic problems ordinarily involved in addressing questions to an indexing and retrieval system. The powerful specificity of citation linkages enables the *Science Citation Index* and ASCA services to disclose efficiently those items which should be of interest even from voluminous peripheral journals not often publishing items "obviously" on a given subject.

The availability of these services now makes practicable several techniques, highly valuable in helping to evaluate research projects from initial conception of an idea through utilization of the results after publication. These techniques also provide new facts for use in evaluating the work of individuals and organizations. Details of the services and some of these applications will be discussed, including charting of the historical development of scientific ideas.

Reprinted from M.C. Yovits, D.M. Gilford, R.H. Wilcox, E. Staveley and H.D. Lemer, Eds., *Research Program Effectiveness, Proceedings of the Conference Sponsored by the Office of Naval Research* Washington, D.C., July 27-29, 1965 (New York: Gordon and Breach, 1966), pp. 135-146.

INTRODUCTION

The use of indirect methods, such as citation analysis, for measuring research performance is not obvious though there is already a sizeable literature in which this has been done on a small scale.⁽¹⁻³⁾

Citation indexing has been extensively developed as a novel method of indexing research literature.⁽⁴⁻⁶⁾ As a result, useful tools for dissemination and retrieval of research information have been made available which provide unique and useful approaches to the literature. The *Science Citation Index* (SCI), printed quarterly and cumulated annually, is primarily used for retrospective searching. *Automatic Subject Citation Alert* (ASCA) is used for selective dissemination of information. ASCA is an individualized weekly alerting service which culls the current literature for answers to specific questions. In both the SCI and ASCA, any known published item, whether an article, book, report, patent, et cetera, which falls within a user's sphere of interest may serve as a question or starting point. Both systems tell the user where any one of a group of question citations has been cited in the footnotes or bibliographies of the more current scientific or technical literature. By the act of citing, the authors and editors of the current literature establish new conceptual relationships between the current work and any earlier item cited. Citation indexing enables one to utilize these newly established relationships in a novel fashion--by coming forward in time from any known published work to more recent related literature. Common publication practices usually also provide some identifying reference numbers within the text of the citing item that facilitate the location of the exact place wherein this conceptual relationship is described or implied. The unambiguous identification of starting question references is relatively simple and does not require the user to rephrase the concepts involved in his question into words, nor to translate those words into standardized indexing terms.

In addition to a general utility for individual scientists, documentalists, and librarians who work with the scientific literature, SCI and ASCA also provide, for administrators, interesting capabilities that can be used in studying, evaluating, and improving the effectiveness of research programs.

REDUCTION OF REPLICATION OF RESEARCH EFFORT

Usually, even as early as the inception of a research idea, the scientist is aware of earlier published material related to his idea, and in fact upon which his idea often is founded. Certainly, by the time an idea has reached the stage of a preliminary project proposal or a grant proposal, there should be appended a bibliography of citations to the pertinent previous literature. If items related to or describing this same "new" idea have already been published, there is a high probability that at least some of the known related literature will have been cited. Any *one* citation to known pertinent literature will alert the researcher and provide an opportunity to see if the idea has indeed already been developed or if modifications in the proposed research are called for.

Through citation analysis, numerous actual cases of unwitting repetition in the publication of research previously recorded by other groups have been established and examined. Even though the Institute for Scientific Information has amassed considerable citation data, finding such examples in the literature is not easy. For every such case of duplication in the literature there probably exist many more instances where the redundancy was discovered after conception of the research idea but in time to abort duplicate publication. Published replications by different groups, and published apologies, are merely the most extreme examples that highlight the inadequacies of the classical literature retrieval capabilities as applied by authors, their colleagues, review board, journal editorial staff, and referees. With the availability of SCI and ASCA, simple checks can be made on the possibility of pre-existing literature at any time throughout the research study and without requiring an established nomenclature in what as yet may be an ill-defined field.

AN OPPORTUNITY TO EXAMINE COMMENTS, CRITIQUES, CORRECTIONS, APPLICATIONS, AND EXTENSIONS OF PARTICULAR WORKS OF AN AUTHOR AS NOTED IN SUBSEQUENT WORKS BY THE AUTHOR'S PEERS

By means of the principle of citation indexing, those responsible

for the administration and evaluation of research programs can uncover current items which cite published works of interest. The administrator can, by careful examination of these citing works, gain the insight and perspective provided by other researchers active in a field and citing the work under consideration. This new information should be considered judiciously in combination with the facts available from all other sources of information ordinarily considered when evaluating researchers or their publications. Until a great deal more research has been conducted into the many sociological ramifications of citation practices, and unless he is fully cognizant of the detailed characteristics of a particular citation index file, the administrator must eschew any naive and most likely erroneous interpretations of the bare statistical counts of the number of citations to particular works.⁽³⁾ The Institute is on record as having cautioned against the promiscuous use of quantitative citation data. However, this by no means implies that such evaluations are not possible.

A preliminary experiment was performed to determine whether or not even the most outstanding scientists differ appreciably from the average in the count of citations to their published works. The file studied was the cumulated *1961 Science Citation Index*, a list arranged by reference (cited) author and with the characteristics given in Table 1. The names of the persons awarded the Nobel Prize in physics, chemistry, and medicine for the years 1962 and 1963 were checked in the 1961 SCI. Statistics on the number of reference citations were gathered and the counts were compared with the values for average authors in this same file as shown in Table 2.

Nobel Prize winners had a significantly higher number of their papers cited, in part reflecting their high rate of publication. However, there was also a higher number of references to each cited work, as compared with the average. The combination of both terms, that is number of items cited and number of reference citations, gives the number of citations per cited author. This term, the impact factor, reveals the greatest difference between groups. There are 30 times as many citations per average prize winner as there are per average cited author in the entire file. Not only is this average for the prize winners higher, but all but one of the 13 individual prize winners had a value at least severalfold greater than that of the average author. The proportion of self-citations was not unusual for the prize winners. The occurrence of 20 or more citations to individual papers was higher for the prize winners as a

group but did not occur in enough instances to make this a reliable statistic.

It should be noted that these calculations make no attempt to correct for possible homographs, that is, a multiplicity of authors that might have the same last name and initials. Nor does this analysis check the number of citations to works in which the individuals concerned appeared as other than first author. Assuming no special propensity for homographs among the names of the prize winners, there appears to be no special trend among these authors to put their names as secondary authors. If such a trend exists, it is unable to mask the unusually large number of citations to the works of these authors. There exists the possibility that, conversely, the prize winners are more likely to put their names first on their papers thus helping to accrue a larger than average number of citations under their names.

Eliminating the one prize winner who had only four citations to his name leaves 37 citations as the next smallest number among the prize winners. Since there were 4,792 reference authors (from among the 257,900 total cited reference authors in the file) who had ≥ 37 citations to their names, we may state that 12 of the 13 prize winners fall in the group of reference authors constituting the top 1.85% of the file. Similarly, if we eliminate the prize winner who had 37 citations to his name we see that the next lowest value is 84 citations. There are 477 reference authors in the entire file with 84 or more citations to their names; therefore, 11 of the 13 prize winners are in the top 0.185% of the file. Thus, in spite of the assumptions and possible flaws in this analysis, it appears rather obvious that a group of scientists, acknowledged by receipt of the Nobel Prize as outstanding, are similarly outstanding when measured by citation analysis. There is no chance here that the receipt per se of the Nobel Prize by these people influenced the number of citations to their works since we have examined the citations which appeared in 1961 for persons who *later* received the prize in 1962 and 1963.

CONVENIENT WAY TO KEEP TRACK OF THE APPEARANCE OF VARIOUS CURRENT WORKS PUBLISHED BY AN INDIVIDUAL OR ORGANIZATION FOR USE BY THE ADMINISTRATION AND PUBLIC RELATIONS DEPARTMENTS

In January 1965 an experimental corporate index was begun in

the SCI and a little later made available in ASCA. All institutions, universities, hospitals, corporations, or other organizations acknowledged as the place of origin of a published journal item are processed as indexing terms. Since the SCI and ASCA processing of journals includes every article, review, technical note, letter to the editor, correction, proceedings of meetings, etc., appearing in the journal, that is, everything other than ads and the most ephemeral news notices, the corporate indexing feature provides a simple means of locating all the published writings, in a specified group of journals, attributed to an organization. The administrator can use this feature as an external source of information confirming the appearance in print of works written in the administrator's own organization and can also be used to follow the papers published by other groups.

Similarly, the administrator can easily identify all of the current articles, letters to the editor, etc., authored by any individual. By using the source index that accompanies the SCI or by entering a source author question into ASCA, the system will identify all the appropriate works no matter where the individual's name appears in the authorship--whether as primary author or otherwise.

Similar data based on patent assignments can be obtained from the 1964 and 1965 SCI Source Indexes which enable one to determine a man's current publications, including all U.S. Patents, and all patents assigned to a specific corporate body.

AN OPPORTUNITY TO STUDY THE UTILIZATION OF THE ITEMS PUBLISHED IN A GIVEN JOURNAL AND THE AMOUNT OF CROSS-DISCIPLINARY INTEREST IN THOSE ITEMS

For those administrators engaged in evaluating the effectiveness of specific publications, SCI and ASCA can provide the data required for such evaluations. The items published in a specific journal, for example, can be followed through citation indexing; and the number and variety of citing publications can be examined. When adjusted for the total number of items published by a journal in a given time period, these statistics can yield individual journal impact factors.⁽⁷⁾ Similarly, the amount of intra- and interdisciplinary utilization and cross-stimulation can be surveyed. There is also provided quantitative evidence of the utilization by journals published in other countries.

AN OPPORTUNITY TO USE AND EVALUATE NEW ALGORITHMIC TECHNIQUE IN WRITING THE HISTORY OF SCIENCE

One of the most elementary applications of citation data, and the derivative SCI and ASCA services, in writing the history of science is that of constructing the introductory section to a current manuscript. Earlier relevant works are easily traced back for many generations by examining the bibliographies of more current known works. The forward-reaching dimension added by SCI and ASCA greatly facilitates the completion of this process by identifying the most recent descendents of the relevant papers uncovered in each generation level. However, this is only one of many reasons for using the citation index for retrieval of information. As will be seen, however, it is in fact difficult to separate completely the historical and bibliographical relationships between scientific events. At the Institute for Scientific Information we have been studying these relationships and the possible application of citation indexing and analysis to the history of science. Essentially what we are investigating is the possibility of developing algorithmic procedures that would match the intellectual results obtained by the historian of science. In a recent experiment,⁽⁸⁾ we tested the heuristic utility of citation indexes in writing the history of science. In this approach, the history of science is regarded as a chronological sequence of events in which each new discovery is dependent upon earlier discoveries. Models of history were constructed consisting of chronological maps or topological network diagrams. Two such models were used. The first is based on the events in the history of DNA as described by Dr. Isaac Asimov in *The Genetic Code*.⁽⁹⁾ The second is based on the bibliographic citation data contained in the documents which are the original published studies of events represented in the Asimov book. The interdependencies of linkages among 40 major events (nodes) included in both network diagrams were mapped and compared. The study confirmed 65% (28 of 43) of the historical dependencies in the Asimov network by corresponding linkages established by citations. However, 31 citation connections were found which had not been noted in *The Genetic Code*. Thus, each model turned up data not found in the other. The analyses, supported by numerous statistical tables and specially constructed citation indexes, show that the original hypothesis is reasonable. The techniques evolved in this study appear useful in writing histories of science by helping to identify key events, their chronology, their interrelationships, and their relative importance.

In conclusion, there are a number of different methods and occasions for the use of citation data in evaluating and improving the effectiveness of research. The more obvious methods involve the use of citation indexes to retrieve pertinent information on research projects while it is sufficiently current to prevent replication of effort. The less obvious methods involve the use of citation indexes to evaluate individual papers and also to use citation counts for determination of impact factors. Finally, citation data can be used, algorithmically, in the reconstruction of the historical development of individual fields or branches of science.

TABLE 1
Statistical Characteristics of
1961 Science Citation Index

Characteristics	Number
Source journal titles processed	613
Source journal issues processed	5,467
Source items	101,944
Total references cited	1,395,530
Anonymous reference items	20,200
Unique reference authors cited	257,900
Different reference items cited	869,900
Cited reference authors with ≥ 37 citations to the sum of their cited works	4,792
Cited reference authors with ≥ 84 citations to the sum of their cited works	477

TABLE 2
Citation Characteristics of Nobel Prize Winners

	First Author	No. Items Cited/Author	No. Citations Per Author	% Self-Citations	No. Citations Per Item	No. Items ≥ 20 Citations Per Author	
<u>Average:</u>							
	1961 File	3.37	5.51	7.8	1.57	0.0039	
	Prize Winners	58.10	169.00	10.5	2.90	0.85	
<u>Nobel Prizes, 1962</u>							
	Physics	Landau, LD	113.00	177.00	-	1.56	-
	Chemistry	Kendrew, JC	26.00	95.00	1.0	3.50	1.00
		Perutz, MF	20.00	84.00	1.2	4.20	1.00
	Medicine	Crick, FH	48.00	119.00	0.8	2.50	-
		Watson, JD	19.00	112.00	-	5.90	3.00
<u>Nobel Prizes, 1963</u>							
	Physics	Wigner, EP	67.00	103.00	2.0	1.54	-
		Mayer, M	35.00	37.00	8.0	1.05	-
		Jensen, JH	3.00	4.00	-	1.33	-

	First Author	No. Items Cited/Author	No. Citations Per Author	% Self-Citations	No. Citations Per Item	No. Items ≥ 20 Citation Per Author
<u>Nobel Prizes, 1963</u>						
(cont.)						
Chemistry	Ziegler, C.	112.00	166.00	-	1.48	-
	Natta, G.	209.00	380.00	26.0	1.82	-
Medicine	Huxley, AF	22.00	109.00	4.0	4.96	-
	Eccles, JC	119.00	422.00	29.0	3.55	1.00
	Hodgkins, AL	65.00	388.00	-	5.96	5.00

REFERENCES

1. Westbrook, J.H., "Identifying Significant Research," *Science*, 132, 1229 (1960).
2. Hodge, M.H., "Rate Your Company's Research Productivity," *Harvard Business Review*, 41(6), 109 (1963).
3. Garfield, E., "Citation Indexes in Sociological and Historical Research," *American Documentation*, 14(4), 289 (1963).
4. Garfield, E., "Citation Indexes for Science," *Science*, 122, 108 (1955).
5. Martyn, J., "An Examination of Citation Indexes," *Aslib Proceedings*, 17(6), 184 (1965).
6. Garfield, E., "Science Citation Index—A New Dimension in Indexing," *Science*, 144, 649 (1964).
7. Garfield, E. and Sher, I.H., "New Factors in the Evaluation of Scientific Literature through Citation Indexing," *American Documentation*, 14(3), 195 (1963).
8. Garfield, E., Sher, I.H., and Torpie, R.J., *The Use of Citation Data in Writing the History of Science*, Philadelphia: Institute for Scientific Information (1964).
9. Asimov, I., *The Genetic Code*, New York: New American Library (1963).

ADDITIONAL READINGS

- * P.H. Abelson, "Coping with the Information Explosion," Editorial, *Science* **154**(3745), 75 (1966); see also A.T. Waterman, "Early Information Evaluation," *Science* **155**(3761), 398, 400 (1967).
A.E. Bayer and J. Folger, "Some Correlates of a Citation Measure of Productivity in Sciences," *Sociology of Education* **39**(4), 382-390 (1966).
- * A.E. Cawkell, "Using References to Retrieve Current Articles," *Radio and Electronic Engineer* **35**(6), 352-353 (1968).
S. Cole and J.R. Cole, "Scientific Output and Recognition: A Study in the Operation of the Reward System in Science," *American Sociological Review* **32**(3), 377-390 (1967). In this paper, the authors state that "The invention of the *Science Citation Index (SCI)* a few years ago provides a new tool which yields a reliable and valid measure of the significance of individual scientists' contributions in certain fields of science".
J.A. Creager and L.R. Harmon, "On-the-job Validation of Selection Variables," Technical Report No. 26, National Academy of Sciences-National Research Council, Washington, D.C., April 15, 1966, 70 pp.
- * E. Garfield, "Citation Indexing for Studying Science," *Nature* (in Press).
E. Garfield, "Selective Dissemination and Retrieval of Information in Biomedical Engineering." Presented at the 18th Annual Conference on Engineering in Medicine and Biology, Philadelphia, Pa., November 12, 1965.
- * E. Garfield, "Patent Citation Indexing and the Notions of Novelty, Similarity, and Relevance," *Journal of Chemical Documentation* **6**(2), 63-65 (1966).
- * M.V. Malin, "The Science Citation Index: A New Concept in Indexing," *Library Trends* **16**(3), 374-387 (1968).
- * J. Margolis, "Citation Indexing and Evaluation of Scientific Papers," *Science* **155**(3767), 1213-1219 (1967);
J.P. Martino, "Research Evaluation Through Citation Indexing," *AFOSR Research*, AD 659 366, D. Taylor, Ed., Arlington, Va., USAF Office of Aerospace Research, 1967, 226-227.
S.J. Routh, "Citation Indexing and Title Permutation," *Changing concepts in librarianship; Proceedings of the 14th Biennial Conference, Library Association of Australia, Brisbane, 1967* (Brisbane: 14th Biennial Conference Committee, 1968).

*Free reprints are available from

ISI® **Institute for Scientific Information**
325 Chestnut Street, Philadelphia, Pennsylvania 19106