

Collection Evaluation and Development Using Citation Analysis Techniques

Elizabeth Fuseler-McDowell
Science and Technology Department
Colorado State University Libraries
Fort Collins, Colorado

ABSTRACT

A description of the various ways in which citation analysis can be used in collection evaluation and development, including a historical overview and current methods. The paper will include methods which can be used by individual libraries to determine the list of "core journals" specific to the individual library.

INTRODUCTION

Developing the most effective library is a basic management issue that concerns librarians in all the various types of libraries worldwide (Pravdic & Oluic-Yukovic). In today's world of rapidly changing and increasing knowledge, periodicals provide users with the most up-to-date information and, therefore, are especially important to researchers and serious students. An increasing growth in publication has resulted in increasing demand for more resources and increasing financial pressure on the limited resources of libraries. In order to operate effectively, librarians must identify literatures of high utility to patrons, and acquire and organize these in such a way as to optimize their usefulness.

Libraries exist to serve specific user groups. It is generally accepted that – to manage the collection effectively – the library's goals and objects, and/or the mission of the parent institution, must be taken into account (Bonn). Because special library collections should be relevant to their specific audiences, and not a national audience, Line (1979) believes that "No measure of journal use other than one derived from a local-use study is of any significant practical value to libraries."

There are some basic premises which underlie all attempts at collection development in that those attempts will

- 1) Gather data for better collection management decisions;
- 2) Assess how well collection meets user needs;
- 3) Act as a check on the implementation of collection development policies;
- 4) Review performance of current collection management policies; and
- 5) Reduce the inherent subjectivity of the selection/deselection process. (Faigel)

One of the primary functions of the librarian in collection development and management is to be a filter between the user and available information (Bradford state of information – "quantitative untidiness of scientific documentation") (Garvey). This intelligent filtering requires that there is some scientific and objective method of filtering (Adewole).

Mosher, Bonn and Faigel have written comprehensive reviews of the existing model of collection evaluation. Kraft has reviewed the literature on collection development models.

The usual methods of evaluating collections are client-centered or collection-centered. Client-centered methods focus on the way the user uses the collection and user satisfaction. These methods look at whether the collection is used and how often. The individual user is the unit of analysis. Techniques used in client-centered evaluations include

- 1) User opinions in which the library's users and librarians are surveyed for the reactions to the collection. This assumes that those surveyed have a frame of reference from which to judge the collection.
- 2) Observations which record the number of items which circulate or number of items which are taken off the shelves and use that number as a measure of the use of the collection. This assumes that high levels of use indicate satisfaction with the collection.

Client-centered methods center on user reaction to the collection as it exists.

For the most part, collection-centered methods attempt to determine the minimum size that will produce a useful collection. The unit of analysis is the collection. Techniques involve measuring the collection against some external standard. That standard may be based on other collections, standardized bibliographies, citation frequency or citation analysis. These methods assume that

- 1) A high number of citations, or listing in a particular bibliography, is an objective measure of a title's worth;
- 2) The items cited are items used in the field of interest of the particular library.

Loertscher has proposed that one must study the collection by looking at the individual subject areas. He proposes that there is no effective method for looking at entire, broad-ranging collections. This is especially applicable in the special library which supports areas of research.

This paper reviews some of the citation analysis techniques which have been used as an objective way of gathering data to assess how well the current collection meets user needs, and to develop the collection.

Introduction to Citation Analysis in Collection Management

Citation use studies establish a source list of books and/or articles based on their inclusion in "core" works. There is an underlying assumption that items important enough to be cited should be in collections. In contrast to checking lists developed by someone's judgement of what is valuable to a field, this "core" list is derived from items actually used by authors. This method allows the literature of an area to define its own characteristics. An advantage to using citations is that they are an unobtrusive way of providing numerical data that can be manipulated (Hockings). Broadus has reviewed the uses of citation counts to evaluate papers, measure scholarly eminence, indicate subject content, and measure library demand.

In 1927 Gross and Gross suggested that citations to a journal in a variety of articles within a subject area are a good indicator of the use of that journal in the library. This was confirmed by Elizabeth Pan in her 1976 dissertation on citation analysis in biomedical libraries. Studies by Broadus have not confirmed the Gross and Gross speculation.

Applications of Bibliographic Citation Analysis to Collection Management

Citation studies “clearly have value, and by offering insights on patterns of information transfer and scholarly communication they have the potential to tell us something about library use and user patterns” (Faigel).

Lopez Method: The Lopez Method or the Citation Technique of Collection Evaluation procedures (CTCE) is a measure of collection depth. Citations are selected at random from a critical bibliography. These are checked against library holdings. This is done at four levels of citation, so that a second set of randomized citations is developed from the critical bibliography. Likewise, a third set is developed from the second, and a fourth set is developed from the third. Each citation is then assigned a value by source level, i.e., ten for level 1, twenty for level 2, etc.

Reworking the method in 1983, Lopez used 100 titles to start, and the checking took place over an entire semester, with rechecks for items not found. Scoring was reversed — eighty for the first level, twenty for the 2nd, etc. He thinks that this works particularly well for new disciplines, where there may not be as many “levels” of publication.

Nisonger repeated the Lopez study twice, each time with a sample size of 25. He tried variations in text, which yielded inconsistencies because the sample was too small and resulted in statistical problems. He did discover that items with recent imprints were more likely to be in their collection. However, this was attributed to recent changes in collection policies rather than the literature.

Mosher used a combination of techniques which included 1) checking a sample chosen from an important subject bibliography; 2) checking the citations in a review article or monograph on the “cutting edge” of research; 3) careful search of all citations in a significant subject bibliography in areas of weakness determined by 1 and 2; and 4) checks of basic lists of most used titles.

Pravdic and Oluc-Yukovic have developed a model which uses multiple factors. They include data sources of international origin, looking at types, quality, and quantity. Type I is *Journal Citation Reports (JCR)*. Type II are journals covered in selective abstracting and indexing (A&I) services and/or databases and/or arbitrarily taken from a fraction of ranked lists of journals in a comprehensive A&I service such as CAS. Type III comes from library holdings of leading universities or national/regional periodical centers. Type IV is the list of journals for the area in *Ulrich's*. They see the quality of data as decreasing from Type I to Type IV.

Citation Count Methods

Most studies which measure citations to specific journals have been done to measure the importance of journals to discipline. There has not been an emphasis on using citations to measure the usefulness of specific journal titles to a specific collection.

Adewole used all citations to 114 livestock journals to derive a Bradford-Zipf distribution. A graphic display of this distribution showed that 18 journals represented 32.3% of the total citations from livestock journals and were determined to be the minimum needed for a usable collection. Subsequent zones were formed from his point. Sixty three percent of the titles supplied over 90% of the total citation.

Resh, in his analysis of citations in two aquatic entomology textbooks and a bibliography of benthic biology, found that

- 1) The most frequently cited journals accounted for 6.1-7.5% of the citations.
- 2) 3.3-5.3% of the total number of journals produced 50% of the total number of journal references; however, over 50% of the journals were cited only once in each source.

Hafner used the distribution of citations to MeSH subject headings to derive a collection of nursing journals. The distribution conformed to a Yule distribution, with the differences between 15% and 25% coverage being four to nine journal titles and the difference between 90% and 100% being 212 to 691 titles.

Several studies have been based on *JCR* categories. Cawkell looked for patterns of citations and mapped connection strengths between the journal in question and all journals in the *Science Citation Index*. Gordon correlated subjective evaluations of "importance" and the major *JCR* indicators - times cited, impact, immediacy, and number of source items. He found that of the indicators, citations were the most highly correlated with subjective evaluation.

Saye and Griffith, using the results from a larger Medline study, 1) selected journal titles; 2) selected specific articles; and 3) selected citations in articles. Steps 2 and 3 added random element to the process. From this core, basic and "slice" list of journals through iterative process using expert opinion and *JCR* were developed. The number of citations derived from each journal varied to reflect the size of the journal and its importance to the subject literature. Five citations were selected from each starting point. All non-monograph items were eliminated. Problems with this method included some duplication of cites, cites to unpublished works, and incorrect cites.

Some studies have tried to use impact factor to indicate the "importance" of a journal. This figure, because it is proportional to the number of citations per number of articles, is thought by some to allow a better comparison of the journals of different sizes (Koenig). This measure is based on frequency of citations to a journal during a specific time frame, divided by the number of source items in that journal. In the case of the Institute for Scientific Information *JCR* impact factor, the two years immediately after publication are the citing years used. While this limited time span is appropriate for most fields, in which articles receive most of their citations, in fields which have significantly longer citing times, this may not be the best of measures. Yearly fluctuations and growth in pages permit more material available for citation in some journals than in others. To develop an impact factor which takes these variables into account will require sophisticated data analysis. At the present, one has not been developed.

Hirst developed a Discipline Impact Factor (DIF), which is similar to the *JCR* impact factor. The DIF differs only in that it uses citations made by a number of known relevant journals in a given discipline. Boyce and Funk derived a similar measure, "quality weight," for journals within a specific field. This quality weight was derived by dividing the number of citations to bibliography items in a journal by the number of items the journal published during a specific time period.

In a 1978 study Pan found no statistical significance between impact factor and use, but statistically significant relationships between citation and number of articles and use.

Narin in his study which defined 50 biomedical fields proposed a Total Citation Influence Measure. This was computed as

$$\frac{\text{\# of times A is cited by all other journals}}{\text{\# of times A cites other journals}} \times \frac{\text{\# refs}}{\text{\# publ.}} \times \text{\# of publ.}$$

where (all figures are for the same time period, y):

\# of refs = total number of references in A over time period y and,

\# of pubs = total number of articles in A over time period y.

Interestingly, among his findings was the fact that in some fields the journals that authors publish in and those they cite are quite different. Garfield has also found this in some of his studies (1977, 1983). Narin (1976) has also used journal influence maps to graphically show the relationships of journals within a field.

Actual Use of Citation Analysis in Collection Management

Perrault notes that citation analysis is not a widely used technique of collection management. He found less of a tendency to use citation dates in the 1980s than there was in the 1970s. In his 1983 dissertation, Helpin found that only 23% of libraries use citation data to assist in decision making. Those who do not use citation data say it is because other criteria are sufficient, subject coverage of studies is inappropriate, and studies are based on national figures (Budd). Table 1 displays the various ways in which citation data has been used for collection selection and management.

Using Citation Data in Collection Management and Development

Librarians with knowledge of the subject area rarely have any trouble identifying the highly-cited or uncited journals. The difficult question is more focused on which middle-ranked journals are important to their collection (Line and Sandison). In some libraries, there may not be a librarian with a knowledge of the same areas of the collection, and the librarians need guidance to select all journals on those subjects.

Citation behavior and relative frequencies of citation have been shown by Price and Garfield (1979) to be field dependent and field specific. This dependency can be used to develop relationships between journals in a specific field and to show a pattern of use.

Citation analysis does provide an objective way of looking at the collection. It is possible to determine how well the collection meets the current interests of the users and to evaluate which periodicals should be added or dropped by using various techniques.

Last year at IAMS LIC in Miami, I was asked, "How can I apply the methods you used to analyze marine biology journals to evaluate my collection?" The answer is that in my opinion you should not use those same methods to analyze your particular collection. The techniques used in that paper and my earlier one on oceanography journals are meant to be used to study the field, not specific collections. The most effective collection is the one that best meets the needs of the specific researchers at the institution, not some broad category of "universal" researcher. In this paper I have tried to give a review of the types of methods that have been used by others to use citation analysis for collection development and management.

Now I would like to offer a few suggestions on how to get started using citation analysis on your collection. I propose that you use an institutionally derived starting point, which therefore reflects the specific interests of your institution. This starting point could be 1) compiled from faculty/librarian opinion on the essential journal titles or definitive works in the specific subject areas; 2) finding the journal most cited by faculty/researchers in your institution by a) analyzing the references of the theses or dissertations coming from your institution; and/or b) faculty research publications; and/or c) assigned readings. Using one, or a combination, of these starting points should give a better reflection of the study and work going on at your institution.

From this beginning journal you can then use the *JCR* to create a lineage map that show how that journal is related to other journals in the field or in broader or narrower fields (Figure 1).

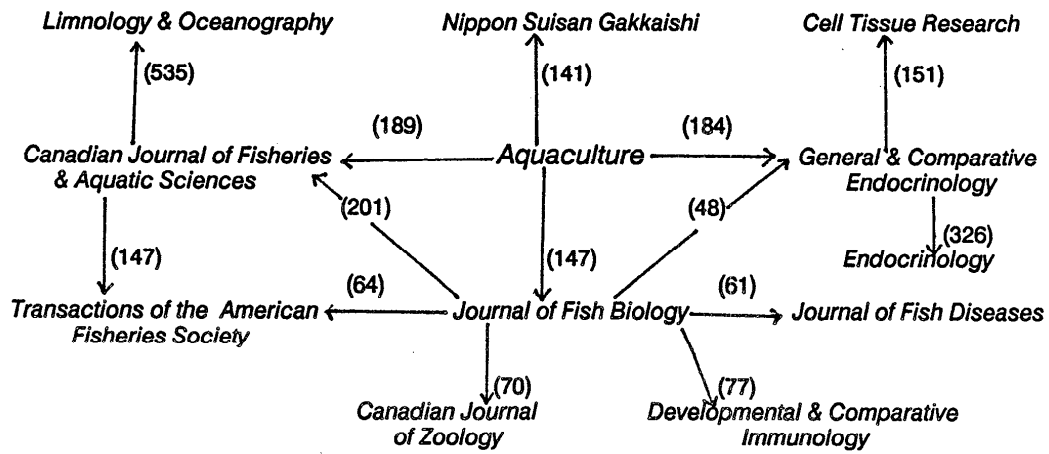
Or, as Resh and Adewole did, you can look at the Bradford distribution of the journals cited in these sources as a measure of how well your collection meets the needs of the users and what titles might be added to the collection or deleted from the collection, if necessary.

Table 1
Comparison of Various Methods of Using Citation Analysis for Collection Management

Collection Level	Scope	Data Source	Analysis	Use for findings	Author(s)
Broad Subject Science	Journals	1) Sampling from subject bibliography 2) Check bibliography of article on "cutting edge" 3) Check items insignificant biblio. in areas of weakness shown by 1 and 2 4) Check lists of most-used & valuable titles	1) - 4) Compare to holdings	Selection	Pravdic
Broad Subject Medicine/ Science	Journals	97 SCT title + 41 librarian suggestions + 31 randomly selected titles. Held in six libraries	Ranked by 1) Citation frequency 2) ISI impact factor 3) Frequency of use in the six libraries	Frequency of use	Pan
Broad Subject Biomedicine	Journals	900 biomedical journals selected	Calculate total citation influence measure (see text)	Subject framework	Narin, et al.
General Subject	Journals	Starting journal selected on basis of obviousness	Calculate discipline impact factor (see text)	Selection	Hirst
General Subject Specific Subject Acoustics	Journals	1) Core journal selected from JCR ranking	Use JCR citation data to rank	Selection	Malin Weinstock Cawkill
Specific Subject Irrigation	Journals	500 random items from Irricab	Check shelves for items	Collection Evaluation	Porta and Lancaster

Specific Subjects 1. Medieval French literature 2. Family therapy 3. American novel 4. Modern British history	Journals	Random sample from critical bibliography	1) References from bibliography sampled 2) Repeat 1 four times 3) Assign weights to four levels in 2 4) Evaluate on basis of possible total	Evaluation	Lopez Nissonger
Specific Subject Livestock	Journals	Citations from 114 current journal subscriptions	1) Rank by number of citations 2) Use Bradford/Zipf graphic distribution to identify core	Identify core collection	Adewole
Specific Subject Nursing	Journals	Citations to MeSH subject headings	1) Rank by number of citations per title 2) Bradford distribution to determine % of coverage	Selection	Hafner
Specific Subject Medical Behavioral Science	Monographs	1) Select 61 journal titles using expert judgement & JCR data 2) Article pool from 1 3) Citations to 239 monographs from 2	Compared to holdings	Evaluation	Saye and Griffith
Specific Subject American literature	Journals Monographs	1) Sample of 1981 MLA Bibliography (253 items) 2) 7149 references from source items	Use Bradford distribution to rank	Management	Budd
Specific Subject Fisheries Acoustics	Journals	1) Select core journal from librarian and researcher opinion	1) Calculate statistical distance 2) Rank by strength of distance	Management	Proposed

Figure 1
Linkage Map for the Journal *Aquaculture*



REFERENCES

- Adewole, Sesun 1987. "Selecting Livestock Periodicals Through Citation Analysis Techniques." *Information Processing and Management* 23:629-38.
- Bonn, George S. 1974. "Evaluation of the Collection." *Library Trends* 22:265-304.
- Broadus, Robert N. 1977. "The Applications of Citation Analyses to Library Collection Building." *Advances in Librarianship* 7:299-335.
- Budd, John. 1986 "A Citation Study of American Literature: Implications for Collection Management." *Collection Management* 8:49-62.
- Cawkell, Anthony E. 1978. "Evaluating Scientific Journals with *Journal Citation Reports* – A Case Study in Acoustics." *JASIS* 29:41-6.
- Faigel, Martin 1985. Methods and Issues in Collection Evaluation Today. *Library Acquisitions: Practice and Theory* 9:21-35.
- Garfield, Eugene 1977. Journal Citation Studies. 20. "Agriculture Journals and the Agriculture Literature." *Essays of an Information Scientist*. 2:272-8, Philadelphia: ISI Press.
- Garfield, Eugene 1979. *Citation Indexing: Its Theory and Application in Science, Technology and Humanities*. New York: Wiley.
- Garfield, Eugene 1983. Journal Citation Studies. 35. "Veterinary Journals: What They Cite and Vice Versa." *Essays of an Information Scientist*. 5:464-72. Philadelphia: ISI Press.
- Garvey, William D. 1979. *Communication: The Essence of Science*. Oxford: Pergamon.
- Gordon, Michael D. 1982. "Citation Ranking Versus Subjective Evaluation in the Determination of Journal Hierarchies in the Social Sciences." *JASIS* 33:55-7.
- Gross, P. L. K. and Gross, E. M. 1927. "College Libraries and Chemical Education. *Science* 66:385-9.
- Hafner, Arthur W. 1976. "Primary Journal Selection Using Citations from an Indexing Service Journal: A Method and Example for Nursing Literature." *Bulletin of the Medical Library Association* 64:392-401.
- Halpin, Jerome Henry, Jr. 1983. *Citation Analysis As the Basis for Journal Collection Management by Academic Libraries* (Ph.D. dissertation). Southern California University.
- Hirst, G. 1974. "Discipline Impact Factors: A Method for Determining Core Journals Lists." *JASIS* 25:131-2.
- Koenig, Michael E. D. 1978. "Citation Analysis for the Arts and Sciences As A Collection Management Tool." *Collection Management* 2:247-61.
- Kraft, Donald H. 1979. "Journal Selection Models: Past and Present." *Collection Management* 3:163-185.
- Line, Maurice B. 1979. "Rank Lists Based on Citation and Library Uses As Indicators of Journal Usage in Individual Libraries." *Collection Management* 3:163-185.
- Line, Maurice B. and Sandison, Alexander. 1975. "Practical Interpretation of Citation and Library Use Studies." *College and Research Libraries* 36:393-6.

Loertscher, David V. 1985. "Collection Mapping: An Evaluation Strategy for Collection Development." *Drexel Library Quarterly* 21:9-21.

Lopez, Manuel D. 1969. "A Guide for Beginning Bibliographers." *Library Resources and Technical Services* 13:462-70.

Lopez, Manuel D. 1983. "The Lopez or Citation Technique of In-depth Collection Evaluation." *College & Research Libraries* 44:251-5.

Malin, Morton V. and Weinstock, Melvin 1975. "Finding the Core Journals of Science Through Citation Feedback." *Bioscience Communication* 1:237-50.

Mosher, Paul H. "Collection Evaluation in Research Libraries: The Search for Quality, Consistency and System in Collection Development." *Library Resources & Technical Services* 23:16-32.

Narin, Francis, Pinski, G. and Gee, H. H. 1976. "Structure of the Biomedical Literature." *JASIS* 27:25-45.

Nisonger, Thomas E. 1980. "An In-depth Collection Evaluation at the University of Manitoba Library: A Test of the Lopez Methods." *Library Resources & Technical Services* 24:329-38.

Pan, Elizabeth L. 1976. *Citation and Use Patterns of Scientific Journals in Biomedical Libraries*. (Ph.D. dissertation) New Brunswick, NJ: Rutgers University.

Pan, Elizabeth 1978. "Journal Citation As a Predictor of Journal Usage in Libraries." *Collection Management* 2:29-38.

Perrault, Anna H. 1983. "Humanities Collection Management — An Impressionistic/realistic/optimistic Appraisal of the State of the Art." *Collection Management* 5:1-23.

Porta, Maria A. and Lancaster, F. W. 1988. "Evaluation of a Scholarly Collection in a Specific Subject Area by Bibliographic Checking: A Comparison of Sources." *Libri* 38:131-7.

Pravdic, Nevenka and Oluic-Vukovic, Vesna. 1987. "Application of Overlapping Technique in Selection of Scientific Journals for a Particular Discipline — Methodological Approach." *Information Processing & Management* 23:25-32.

Price, Derek de Sola 1963. *Little Science, Big Science*. New Haven: Yale University Press.

Resh, Vincent H. 1985. "Periodical Citations in Aquatic Entomology and Freshwater Benthic Biology." *Freshwater Biology* 15:757-66.

Saye, Jerry D. and Griffith, Belver C. 1988. "Monograph Support Provided the National Library of Medicine and Its Regional Medical Libraries in the Medical Behavioral Sciences." *Bulletin of the Medical Library Association* 76:295-305.