DOCUMENT RESUME

ED 134 970 CS 003 239

AUTHOR Mayes, Bea
TITLE Distributions of Journal Citations in Small Collections of Reading Research.
PUB DATE [76]
NOTE 18p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Bibliographic Citations; Bibliographies; Information Retrieval; *Library Collections; *Periodicals; *Reading Research; *Search Strategies

ABSTRACT

The distribution of reading-research citations was investigated in three populations of journals. The rule of Pareto-like distribution was confirmed as appropriate for determining the number of journals that would contribute half the citations in populations of 26 to 112 journals. In populations of 42 to 112 journals, 24% to 29% of the high-ranking journals were necessary to account for 80% of the citations. In two further samples, chosen on different criteria, more than 35% of the journals were necessary to account for 80% of the citations. Implications for both researchers and librarians are discussed. (Author/AA)

Bea Mayes

Distributions of Journal Citations in Small Collections of Reading Research

AN ABSTRACT

Growth rules were applied to citation distribution in three small journal populations derived from collections of reading research. The rule of Pareto-like distribution was confirmed as appropriate for determining the small number of journals that would contribute half the citations in populations of from 26 to 112 journals. In journal populations of from 42 to 112 journals, which represented entire collections or entire random samples, from 24% to 29% of the high ranking journals were necessary to account for an 80% core of citations. In the unique portions of two random samples, more than 35% of the journals were necessary to account for an 80% core of citations.

Distributions of Journal Citations in Small Collections of Reading Research

Bea Mayes

In any area of inquiry rules of growth and distribution of population are useful to forecasters and practitioners. Price (1963) was interested in the growth of science. He applied rules of growth to scientific contribution. The growth of cities and income distribution had been shown to follow Pareto-like distribution. Price demonstrated that scientific productivity and journal usage also followed a Pareto-like distribution. As a population grows in Pareto-like distribution, the large producers become larger, while the smaller producers merely increase in numbers. As the population of all cities grows, the larger cities become still larger, while the ranks of the small cities are added to by the development of still more small cities. Similarly, the high producing scientists produce a large share of all scientific papers, while the ranks of low producers are swelled by scientists who may publish only one paper. In a bibliography, a small number of journals will answer one half the user demand, while all the other journals together will satisfy the remainder of the requests. The dividing line between the high producers in any of these populations and the low producers is the number equal to the square root of the total population, according to the rule of Pareto-like distribution. This rule defines the small number of large producers and separates them from the many small producers within the population.

Price confirmed Pareto-like distribution using data collected by Urquhart (1958) at the Science Library in London. In this large scientific collection he also found that less than 10% of the journals were sufficient to account for an 80% core of user demand.

The notion of rules of distribution of citations or demand in a bibliographic collection is of particular interest to librarians and scholars. The researcher reasoned that the task of making distribution information useful was two-fold. First, these rules must be shown to be generally applicable. Second, known collections of research must be analyzed to discover the sources most pertinent to given topics. Once the major journal contributors on a topic have been identified, both librarians and students will find their jobs simplified. The librarian organizing a collection can depend on distribution analyses in known collections to determine the most cited journals on a given topic. Students looking for a good central core of information on a topic need only consult the relatively small number of journals found to be large contributors on that topic. The application of general rules concerning use and citation distribution in collections will allow librarians and students in a field to identify those journal sources which will be most fruitful in studying research on a given topic.

The Study

This study developed from Price's notion of growth rules. It applied his numerical results to bibliographic citations

in sample populations from two bibliographies of reading research in order to test the wider application of the growth rules. The investigator contended that in small journal populations, Price's finding that 10% of the journals in a bibliography could supply an 80% core of the products was incompatible with the rule of Pareto-like distribution. If Price's numerical findings were true growth measures they could be applied to other populations. The numerical relationships would apply equally to citations or demand. This study tested these rules by applying them to citation distribution in small journal populations.

Two bibliographies on closely related topics were used, the Literature Search in Reading bibliography compiled by Kling and his associates (Davis, 1971) and the ERIC/CRIER bibliography Recent Research in Reading (CCM Information Corporation, 1970), both collections of reading research. Journal populations derived from these two bibliographies were used to test the rule of Pareto-like distribution and Price's findings concerning an 80% core of a bibliography. Journal citations accounted for a major portion of the source citations in the sample populations studied.

Background of the Study

Price found Pareto-like distributions in scientists' productivity and in the requests for journals in a large scientific bibliography. In addition, his reanalysis of Urquhart's data from the Science Library in London showed

that in this large scientific collection 10% of the top ranked journals were sufficient to meet 80% of the user requests. From an arithmetical point of view, Price's two measures were incompatible in small bibliographies. For example, if a collection contained 100 journals, the ten top-ranked journals would account for 50% of the products, if one used the square root rule of Pareto-like distribution. However, according to Price's results from the Science Library, these same ten journals should also account for a core of 80% of the products. Furthermore, in a population of 49 journals, seven journals would account for 50% of the products according to the rule of Pareto-like distribution, while only 5 journals would contribute a core of 80% of the products if one applied Price's results. In a population of nine journals, three journals would supply 50% of the products according to Pareto-like distribution, while only one journal would contribute an 80% core of the products according to Price's findings. These are incompatible conclusions. As the size of a collection decreases numerically below 100 the top 10% of the journals which is said to produce 80% of the products becomes numerically smaller than the square root of the total which should account for 50% of the products. This results in increasing incompatability.

It seemed apparent that either the rule of Pareto-like distribution or Price's measure for determining an 80% core in a bibliography was inappropriate in small journal populations. The rule of Pareto-like distribution is appro-

priate to other populations (Price, 1963). For this reason it seemed likely that this rule would be confirmed in the journal populations tested. Price's rule for determining an 80% core was determined from the products of a very large bibliography. It seemed probable that this rule would not be confirmed in the small journal populations used in this study. The researcher undertook to determine which, if either, of these rules applied to citation distribution in three journal populations derived from collections of reading research.

The hypothesis of this study were:

1) analysis of citation frequencies in the journal populations would confirm that the citation distributions approximated a Pareto-like distribution. That is to say, the number of high ranking journals equal to the square root of the journal population would supply 50% of the citations.
2) analysis of citation frequencies in the journal populations would show that 10 per cent of the highest ranking journals did <u>not</u> supply 80% of the citations in the small journal populations of the study.

Related Research

Summers (1968) presented a detailed analysis of the journal contributors to the William S. Gray Collection of reading research. Summers found that from 24-29 per cent of his high ranking journals made up an 80 per cent core of citations. The contents of ERIC/CRIER Recent Research in Reading is summarized by Laffey (1970). Mayes (1973) summarized the compilation of the Literature Search in Reading bibliography. The efficacy of titles in describing contents was studied by Bernard (1963) and Sedano (1964). Hammond (1965) discussed the incompatibility of indexing by different agents.

Price (1963) emphasized the exponential growth of science in general, illustrating this trend with figures on growth of journals, productivity of scientists, and utilization of scientific bibliography. A Pareto-like distribution is characteristic of growth in these areas and reflects crystallization in the growth process, with the largest contributors growing at an ever increasing rate. Price demonstrated Pareto-like distribution in scientific journal utilization. He concluded that a scientific bibliography followed a Pareto-like distribution in its growth. Based on Urquhart (1958), he also noted that less than 10 per cent of the available science serials were sufficient to meet 80% of the demand by users.

Procedures

In order to provide comparable data, the study used similar indexes for the Literature Search in Reading bibliography and the ERIC/CRIER bibliography Recent Research in Reading. A KWOC (Key-Word-Out-Of-Context) index was prepared from the ERIC/CRIER materials to match the already existant KWOC index for Literature Search in Reading (A complete source listing was not available for Literature Search in Reading at the time.) The initial sampling came from comparable KWOC (Key-Word-Out-Of-Context) indexes for the two bibliographies. The procedure was experimental. It contrasted unique word universes obtained from random samples taken from the KWOC index for Literature Search in Reading and a similar KWOC index compiled for ERIC/CRIER Recent Research in Reading (Mayes, 1973). The Key-Word-Out-Of-Context (KWOC) index is a permutation index based on titles. Each entry consists of a code, an index word, and the entire title. The computer collects all titles in which a specific word appears and arranges them in the index under that key title word. For example, all titles in which the word "attention" appears are listed under the key word "attention" in the index. The number of citations in the unique universe of a sample consisted of all the unique key words in the sample and each of the titles listed under them. The unique universe of KWOC title words from each index and the journal sources represented formed the main body of data of Mayes' study. A secondary analysis identified all sources represented in the random sample of title words from Recent Research in Reading, as

well.

The journal populations in the study were identified by tracing the titles identified in the samples to the source materials they represented. Three journal populations were analyzed: The journals represented in the unique portion of the random sample from the KWOC index of Literature Search in Reading, the journals represented in the unique portion of the random sample from the KWOC index of Recent Research in Reading, and the journals represented in the entire random sample from the KWOC index of Recent Research in Reading.

Results

Application of Rule of Pareto-like Distribution

The square root rule of Pareto-like distribution was applied to the contributions of journals traced from three sample populations from the KWOC indexes. In the unique portion of the Literature Search sample, 11 of the 112 journals contributed 48% of the 332 journal articles identified. In the unique portion of the ERIC/CRIER sample, 5 of the 26 journals contributed 55% of the 104 journal articles identified. In the entire ERIC/CRIER sample 7 of the 51 journals contributed 65% of the 432 journal articles identified. The specific journals identified in application of the rule of Pareto-like distribution appear in Table 1. The journals identified contribute approximately 50% of all the journals identified in each of the three sample populations. That is to say, the Journal of Experimental Psychology, Journal

of Educational Psychology, the Journal of Verbal Learning and Verbal Behavior, Perception and Psychophysics, Science, Reading Teacher, Child Development, Journal of Educational Research, American Journal of Psychology, Perceptual and Motor Skills, and British Journal of Educational Psychology contribute approximately 50% of the journal citations in the unique portion of the Literature Search in Reading sample.

Perceptual and Motor Skills, Journal of Psychology, Psychological Reports, Journal of Reading, and Journal of Applied Psychology contribute approximately 50% of the journal citations in the unique portion of the Recent Research in Reading sample. (The Journal of Applied Psychology ties in contribution with the Journal of the Reading Specialist. Either may be identified in fifth rank.)

Psychological Reports, Perceptual and Motor Skills, Journal of Psychology, Elementary English, Reading Teacher, Journal of Reading and Journal of General Psychology contribute 65% of the journal citations in the entire sample from ERIC/CRIER Recent Research in Reading.

TABLE 1

APPLICATION OF PARETO-LIKE DISTRIBUTION TO JOURNAL CITATIONS DERIVED FROM THREE SAMPLE POPULATIONS

Literature Search in Reading: Journal citations from unique portion of random sample from KWOC index. These journals supply 48% of the journal articles identified.

Rank	Journal	No. of Articles
1.0	Journal of Experimental Psychology	38
2.0	Journal of Educational Psychology	20
3.0	Journal of Verbal Learning & Verbal Behavior	17
4.5	Perception and Psychophysics	13
4.5	Science	13
6.0	Reading Teacher	12
7.5	Child Development	11
7.5	Journal of Educational Research	11
9.5	American Journal of Psychology	9
9.5	Perceptual and Motor Skills	9
11.0	British Journal of Educational Psychology	7

ERIC/CRIER Recent Research Reading: Journal citations from unique portion of random sample from KWOC index. These journals supply 55% of the journal articles identified.

Rank	Journal	No. of Articles
1.0	Perceptual and Motor Skills	20
2.0	Journal of Psychology	16
3.0	Psychological Reports	12
4.0	Journal of Reading	7
5.5	Journal of Applied Psychology	6
	or	
5.5	Journal of the Reading Specialist	6

TABLE 1

(Continued)

ERIC/CRIER Recent Research in Reading: Journal citations from entire random sample from KWOC index. These journals supply 65% of the journal articles identified.

Rank	Journal	No. of Articles
1.0	Psychological Reports	73
2.0	Perceptual and Motor Skills	61
3.0	Journal of Psychology	40
4.5	Elementary English	32
4.5	Reading Teacher	32
6.0	Journal of Reading	23
7.0	Journal of General Psychology	20

Determination of an 80% core

Price concluded that in a large bibliography less than 10 per cent of the high ranking journals were sufficient to meet 80 per cent of the demand. The contributions of 10 per cent of the high ranking journals were identified in the three sample populations: In the unique portion of the KWOC sample from Literature Search in Reading, 11 journals (10%) supplied 48% of the citations. In the unique portion of the KWOC sample from the ERIC/CRIER bibliography, 3 journals (10%) supplied 46% of the citations. In the ERIC/CRIER total sample, 5 journals (10%) supplied 55% of the KWOC citations.

The high ranking journals contributing 80% of the citations in the three journal populations were identified: In the unique portion of the KWOC sample from Literature Search in Reading, 44% of the journals supplied 81.3% of the citations. In the unique portion of the KWOC sample from the ERIC/CRIER bibliography, 38% of the journals supplied 80.5% of the citations. In the ERIC/CRIER total sample 29% of the journals supplied 80.2% of the KWOC citations.

Discussion and Conclusions

The data of this study confirmed the hypotheses:

1) In the three sample populations analyzed the large journal contributors to each population contributed approximately 50% of the citations. The rule of Pareto-like distribution was confirmed.

2) In the three small sample populations analyzed Price's findings concerning determination of an 80% core of a bibliography was not upheld. Price found that in a large bibliography, less than 10% of the journals satisfied 80% of the demand. Analysis of citations in small bibliographic populations has shown that a considerably greater proportion of journals, 20%-30%, was necessary to account for an 80% core of citations in an entire collection or entire random sample.

Price's rules of distribution were applied to three journal populations: The journals represented in the unique portion of the random sample from the KWOC index of Literature Search in Reading, the journals represented in the unique portion of the random sample from the KWOC index of Recent Research in Reading, and the journals represented in the entire random sample from the KWOC index of Recent Research in Reading.

The results of application of the rule of Pareto-like distribution to the three sample populations derived from the KWOC indexes showed fairly good correspondence with expected results. The square root of the total population supplied 48%, 55% and 65% of the citations in the journal populations studied. The investigator concluded that these samples tended to follow the rule of Pareto-like distribution. The number of high ranking journals equal to the square root of the total population supplied approximately half the citations in the journal

populations analyzed. In the three sample populations studied 44%, 38% and 29% of the top ranked journals were necessary to supply 80% of the citations. The journal populations totaled 112 journals, 26 journals and 51 journals, respectively. The first two populations represented partial samples. The last population represented a complete random sample from the ERIC/CRIER collection. While these bibliographies represented closely related topics, the most cited journals in the two collections differed considerably.

In view of the lack of correspondence concerning the measure for determination of an 80% core of journal citations between the Price-Urquhart data and the results of this study, further exploration is indicated. It seems likely that the growth of citations, or demand, does not proceed as a fixed percentage of the total population. Rather, it seems likely that the curve representing the proportion of journals contributing an 80% core of journal citations from a collection would demonstrate a geometric relationship between the size of the collection and the number of high ranking journals producing an 80% core of citations. Clearly the data in this study showed that for small journal populations ranging in size from 26 to 112 journals, considerably more than 10% of the journals were necessary to supply 80% of the citations. Based on the data of Mayes' study, the data of Summers (1968) and the data of Price (1963), the number of journals contributing the core 80% of the bibliographic products in those collections of

research tended to vary inversely with the size of the population.

The advantages of using of growth facts in connection with bibliographic citations such as those of Pareto-like distribution and core journal analysis accrue to both researchers and librarians. Once a bibliographic sample has been analyzed, researchers interested in a good central portion of available data on a subject, need only consult the number of top-ranked journals equal to the square root of the total population. Conversely, librarians building a collection on a subject can organize a good working collection of representative articles by consulting a relatively small number of journals, once a master bibliography in the subject has been analyzed as to its journal contributors. In any established area, a relatively small number of journals will contribute 50 per cent of the citations. Another, somewhat larger number of journals will contribute an 80% core of citations. Growth tendencies and rules of distribution such as those applied in this study give new and useful information to those interested in building and using journal collections. Thus, the verification of growth tendencies and rules of distribution in bibliographic populations is of importance to both librarians and researchers.

REFERENCES

Bernard, Jessie and Shilling, C.W. "Accuracy of Titles in Describing Content of Biological Sciences Articles", BSCP Communique, 10-63. Washington, D.C.: American Institute of Biological Sciences, 1963.

CCM Information Corporation. Recent Research in Reading. A Bibliography 1966-1969. New York: CCM Information Corporation, 1970.

Davis, Frederick B. (Ed.). The Literature of Research in Reading with Emphasis on Models. Final Report. Targeted Research and Development Program in Reading, Office of Education, National Center for Educational Research and Development, United States Department of Health, Education and Welfare, Washington, D.C.: Contract No. OEC-0-70-4790, 1971.

Hammond, W. "Dimensions in Compatibility". In Newman, S.M. (Ed.), Information Systems Compatibility. Washington; Spartan Books, 1965, 7-17.

Laffey, J.L., "Forward", Recent Research in Reading, A Bibliography, 1966-1969. New York: CCM Information Corporation, 1970, VII-XIII.

Mayes, Beatrice H. "A Comparison of Title Words and Journal Citations from the Literature of Reading". Unpublished doctoral dissertation, Rutgers University, 1973.

Price, Derek J. de Solla. Little Science, Big Science. New York: Columbia University Press, 1963.

Sedano, J.M. "Keyword-In-Context (KWIC) Indexing; Background, Statistical Evaluation, Pros and Cons, and Applications". Thesis, University of Pittsburgh, 1964.

Summers, Edward G. "Reading Research Literature: Identification and Retrieval", Reading Research Quarterly, 4 (1), 1968, 5-48.

Urquhart, D.J. "Use of Scientific Periodicals", International Conference on Scientific Information, National Academy of Sciences-National Research Council, Washington, D.C., 1958.