# Citation Behavior: Classification, Utility, and Location

**V. Cano**
*EBSCO. Europe, P.O. Box 204 1430 AE AALSHEER, The Netherlands*

**This study tested empirically the citation behavior model of Moravcsik and Murugesan and examined the hypothesized relationships between three variables: reported citation type, reported utility level, and citation location. A group of elite scientists constituting an "invisible college" were asked to classify the references they had made in two of their recent papers following the model in question, and to judge the utility content of each reference cited. The response rate constituted 66% of a total of 42 questionnaires. A total of 344 references were examined. Some departures from the Moravcsik and Murugesan citation behavior model were found, as well as indications of complexities of both citation motivation and citation evaluation. Many citations were paired in categories presumed dichotomous by the model: 29 instances of cited documents were reported to have both a conceptual and an operational nature. Indeed, a document may contain many items of information that may be cited for a number of reasons. It is concluded that studies focusing on elements of information cited (coupled to their location parameters) as opposed to full citations, are needed to develop empirically based models reflecting the patterns of information use and the citation behavior of a scientific community.**

## Introduction

Research in scientific communication flow has concentrated on the identification of steps in communication and the role they play in structures of communication. However, little is known about the processes by which recorded discourse is received, assimilated and integrated into a scientist's contribution to new information. Utilization of information is a difficult variable to make operational [1].

The scientific paper, inasmuch as it is a countable end product of research in its final and public form, contains one type of element that has been considered an indicator of the utility content of printed information: the citation to the work of other scientists. Citations have been suggested as the unit for the study of the diffusion and use of information [2].

Since the creation of the Science Citation Index, (SCI), the use of reference citations as primary research tools has proliferated. However, the application of bibliometric variables as primary analytic tools in the history and sociology of science has not been without criticism. The central problem seems to be that, although citations are easily counted, it is not certain what is being measured [3]. In other words, neither citation behavior, nor citation use or citation impact has been satisfactorily explained. The present study addresses the notion of pure citation counts as a simplistic explanatory model for citation behavior patterns and examined the relationships between a theoretical model of citation behavior, the frequency and utility-level of specific citation types and the applicability of a citation location parameter as a novel bibliometric variable.

Numerous theoretical models have been suggested to explain the complexities of citing behavior. Small [4] and more recently Brookes [5] have presented reviews of the literature in citation analysis and citation behavior, respectively. Most of the models are quite similar in nature [4], although each one seems to have evolved independently and has used different terminology to express almost equal citing motivations. In addition, most models have been presented with little or no empirically derived data to support their claims.

Prabha, [6] in an empirical study of citation behavior concluded that less than one third of the sources cited were considered essential by those who cited them. More recently, Brookes [7] has called attention to the lack of empirical work in this area and has provided preliminary data of citing motivations. An empirically based model of citation behavior is needed to differentiate essential from non-essential citations in the production of a scientific paper. One practical application is immediately apparent to library collection developers but, more generally, such a model would help elucidate some of the complexities of the citing phenomenon.

The present study examined the relationships between three variables: citation type (as reported by the research subjects themselves), reported utility level, and citation location. Twenty-one elite scientists working in the field of structural reliability, a subfield of structural engineering, were selected as research subjects. The members of this group were asked to classify the references they had made in two of their recent papers following the citation

Received November 21, 1986; revised February 23, 1987; accepted March 2, 1987.

behavior model and definitions proposed by Moravcsik and Murugesan [8]. This model consists of eight categories arranged into four dichotomous pairs, two of which pertain to relationships between cited and citing documents and the other two to the use made of the cited document by the citer. The model is conceived as mutually exclusive between the pairs but not across the categories [Table 1].

The present study had three purposes: to test the applicability of the model of citation behavior proposed by Moravcsik and Murugesan as a model of information use; to present the frequency and utility level of citation-types appearing in typical technical papers written by elite scientists, and to examine the usefulness of a citation location parameter as a novel bibliometric variable to refine the crudeness of pure citation counts.

## Methodology

The research subjects were selected by three criteria: citation count, publication count and peer evaluation. The research subjects constitute an "invisible college"; its members share common educational links, coauthorships, memberships to distinguished scientific societies and international prizes. The selection of the group did not involve citation or publication count thresholds, but rather the elaboration of a joint productivity/citation index coupled with a peer review evaluation index. Details of these indices are found in Cano [9].

The 21 research subjects were presented with a copy of two of their recent publications attached to a self-administered questionnaire. The publications were technical papers published in two core journals of the area: the ASCE (American Society of Civil Engineers) *Journal of the Engineering Mechanics Division* and the *ASCE Journal of the Structural Division*. The research subjects were instructed to classify their references following the model as well as the definitions of Moravcsik and Murugesan [8]. Specific connotations due to the labeling of the categories in the model were thought to be offset by the clarity of the definitions of each category. Likewise, the mutual exclusivity and dichotomous nature of each pair in Table 1 was not mentioned specifically, but it should have been clear to the research subjects through the definitions.

In addition, the research subjects were asked to judge the utility content of each reference they had cited. They were asked to rate on a four point scale the level to which each reference had contributed to the production of the technical paper. The utility content of a reference was therefore defined as that which makes a reference indispensable in the production of novel information. This definition was considered appropriate because the purpose of the study was to present the frequency with which specific citation-types appear in a technical paper written by an elite scientist.

A procedure aimed at locating concentrations of citations within the set of papers analyzed was performed. Location parameters for each citation within each paper were set by measuring the distance from the beginning of the paper to the location of each citation within the paper. The location was then expressed relatively, as a fraction of the total paper length. A standard template divided into tenths provided the coordinates to trace citations within pages. This procedure is easily understood by imagining a hypothetical paper being typed as a long continuous file on a scroll, the type-size and line length being adjusted so that the article has unit length. The location parameter is thus equal to the normalized distance of the citation from the top of the scroll.

TABLE 1. Moravcsik and Murugesan Model.

| | | |
|---|---|---|
| (1) Perfunctory | vs. | (2) Organic |
| (3) Conceptual | vs. | (4) Operational |
| (5) Evolutionary | vs. | (6) Juxtapositional |
| (7) Confirmative | vs. | (8) Negational |

## Findings

### *Citation Distribution*

The response rate constituted 66% of a total of 42 questionnaires. A total of 344 classified references (not counting multiple mention of citations within a paper) were classified into the eight categories of the model. Since the categories are not mutually exclusive, some references were classified into more than one category by the research subjects. These were counted in as many categories as they were assigned. It should be clarified that Moravcsik and Murugesan tested their model by classifying only theoretical references of over 30 articles in physics. By contrast, the present paper had the authors of the 42 papers examined perform the classification themselves. This difference in methodology might account for the differences in the citation-type frequencies presented by Moravcsik and Murugesan [8] and the frequencies presented in Table 2.

### *Paired Citations*

The research subjects reported multiple use of cited documents. This was in conflict with the model since citations were paired within the presumed dichotomous categories. In particular there were 29 instances of cited documents reported to have both a conceptual (theoretical) and an operational (instrumental) nature. This is a specific

TABLE 2. Citation types.

| Type | % | Total |
|---|---|---|
| (1) Perfunctory | 26 | 122 |
| (2) Organic | 21 | 99 |
| (3) Conceptual | 19 | 91 |
| (4) Operational | 12 | 58 |
| (5) Evolutionary | 14 | 66 |
| (6) Juxtapositional | 4 | 17 |
| (7) Confirmative | 2 | 11 |
| (8) Negational | 2 | 9 |

example of documents reported to have been cited both for theoretical and instrumental elements in them. Table 3.

TABLE 3. Paired citations.

| Type | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 13 | 2 | 0 | 3 | 5 | 3 |
| 2 | | 28 | 24 | 14 | 0 | 2 | 0 |
| 3 | | | 29 | 15 | 1 | 2 | 2 |
| 4 | | | | 11 | 0 | 1 | 0 |
| 5 | | | | | 7 | 3 | 1 |
| 6 | | | | | | 2 | 0 |
| 7 | | | | | | | 0 |

*Utility Level*

On the basis of the data collected, perfunctory and negational citations (first citation type group), were given a low utility level and exhibit similar distribution shapes, Fig. 1(a). However, in spite of the reportedly low utility level exhibited by the perfunctory type, research subjects reported it nonetheless, to be the most frequent category (26%). This can be interpreted as a sign that the research subjects were not unduly affected by any presumed value-ladennesss in the labeling of the categories of the model. On the other hand organic, conceptual, operational and

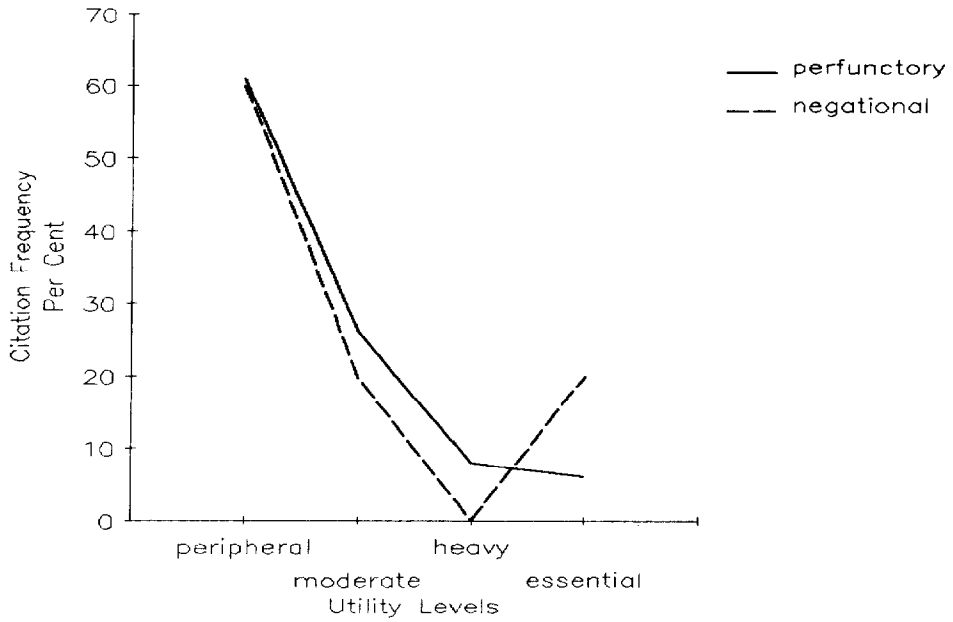


FIG. 1(a). First citation type group.

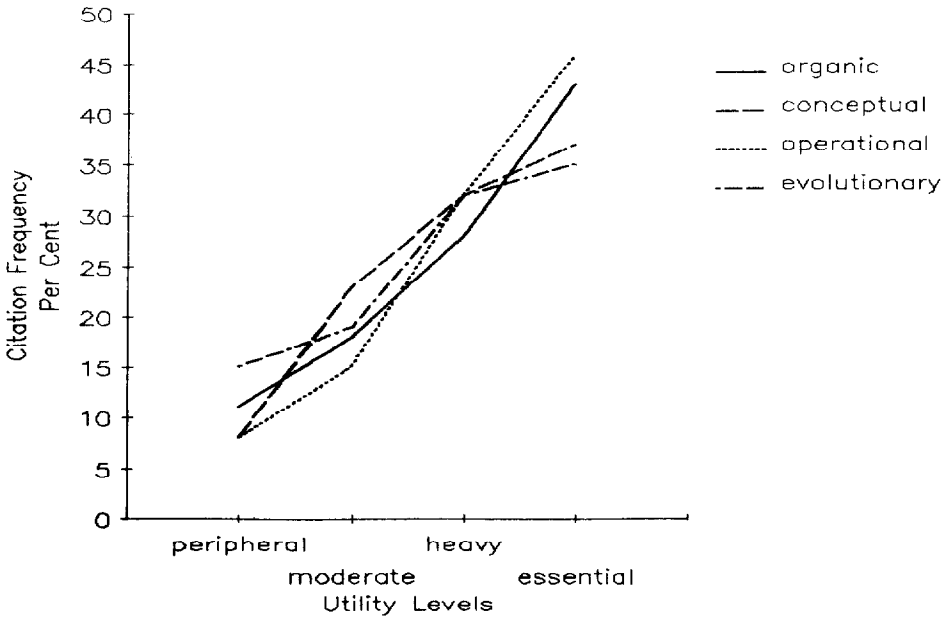


FIG. 1(b). Second citation type group.

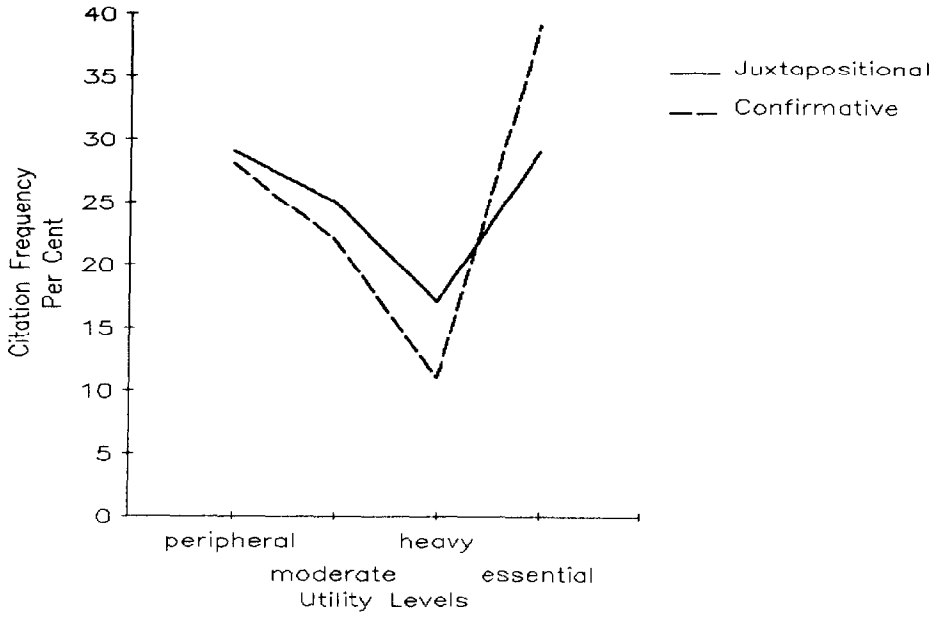


FIG. 1(c). Third citation type group.

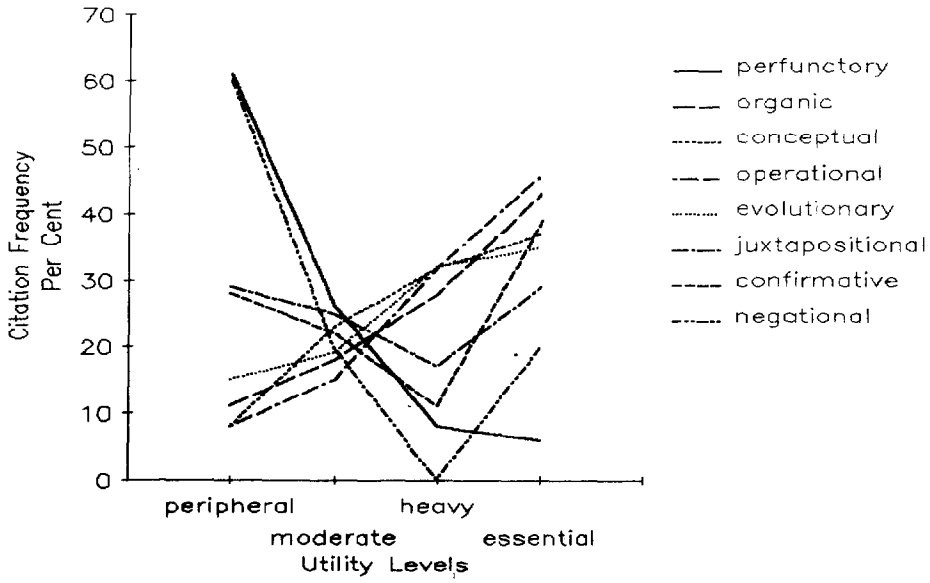


FIG. 1(d). Reported utility level per citation types (all groups).

evolutionary categories (second citation type group) tended to be given a high utility-content level and exhibit similar distribution shapes, Fig. 1(b). Juxtapositional and confirmative categories (third citation type group), were given values fairly uniformly over the four utility levels and also exhibit similar distribution shapes, Table 4 and Fig. 1(c). An overview of the distribution shapes, of the three citation type groups is presented in Fig. 1(d).

The distinction between the three distribution shapes, (citation types over utility levels) is succinctly brought out in the mean levels shown for each type. A chi-square test was used to test the statistical significance of the difference in the distribution of utility levels for the three citation type groups. A chi-square rejection level of 16.812 at a 0.01 confidence level with 6 degrees of freedom was established. The calculated chi-square value of 146.3 established a significant difference between the three distribution shapes.

### *Citation Concentration*

Citations tended to be concentrated in three areas of the paper defined as: beginning section (up to the 15 percentile), middle section (from the 20th to the 75th percentile), and end (from the 80th percentile). The largest concentration was located in the first 15% of the paper; this agrees

with citation densities reported by Voos and Dagaev [10]. Perfunctory citations comprised the largest category of citations in this section (33%). Organic citations (32%), comprise the largest category in the middle section of the paper. This section shows a decline around the twentieth per cent mark of the paper and minor peaks occur near the fortieth and sixtieth per cent. Interestingly, organic citations also comprise the largest category at the end of the paper (41%). Overall, citation frequency steadily declined after the fifteenth per cent location mark. Figs. 2–4.

## Conclusions

Citations seem to be more heavily concentrated in the first 15% of the technical papers studied, corresponding roughly to their introductory sections. However, over one third of the citations in this section were of a perfunctory nature and of very little use to the research subjects in their production of novel information. This result might lead to the hypothesis that citations located at introductory sections of technical papers represent a mere "setting of the stage" and have very little informational utility to the authors of the papers. This does not mean that citations in these sections are worthless in information content to all readers. Merely that to elite scientists citations in introductory sections might present exclusively the background literature in relation to which their new contribution is to be examined.

The hypothesis of citations in introductory sections constituting a "setting" for new research to be examined opens

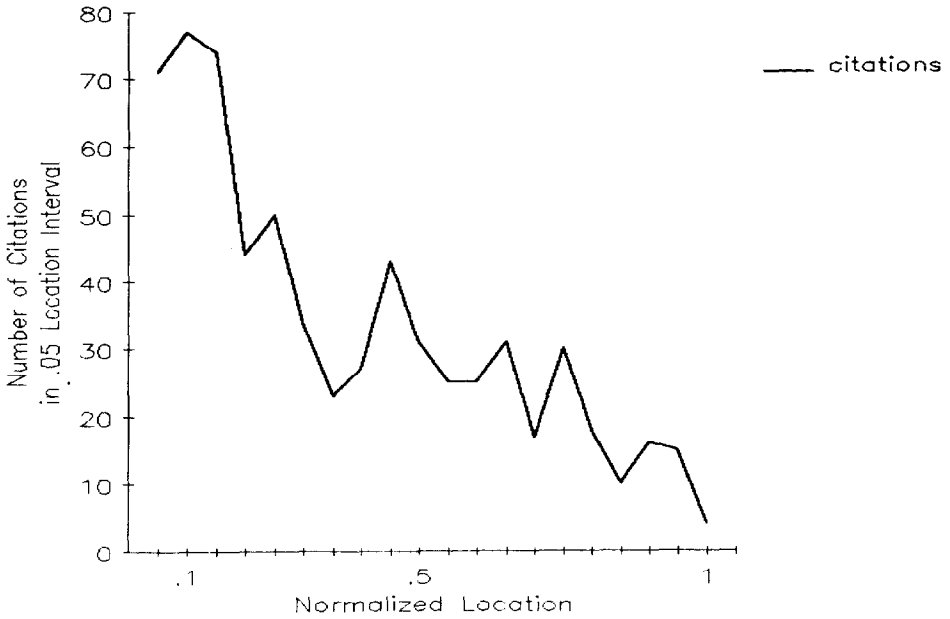


FIG. 2. Citation distribution over normalized location.

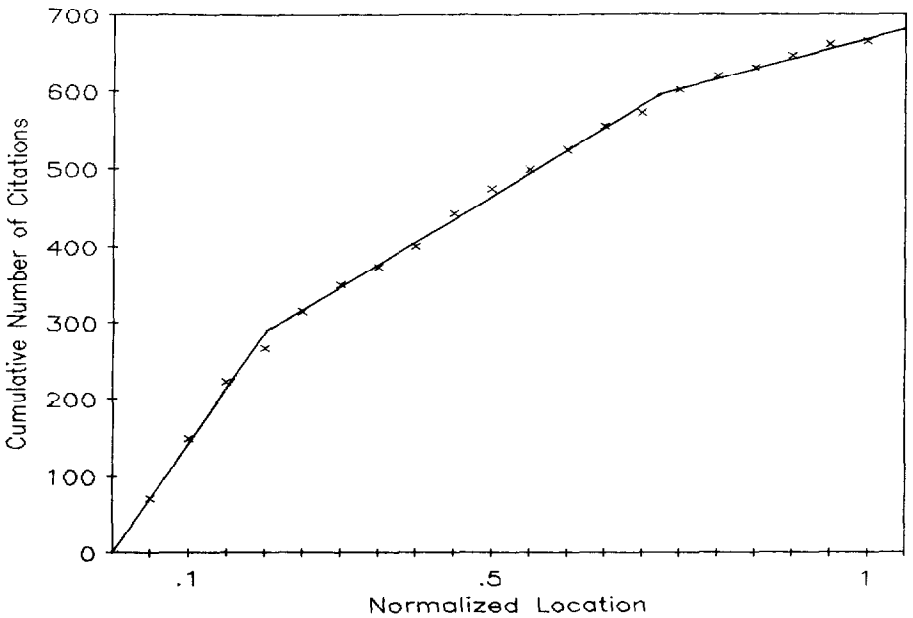


FIG. 3. Cumulative distribution of citations.

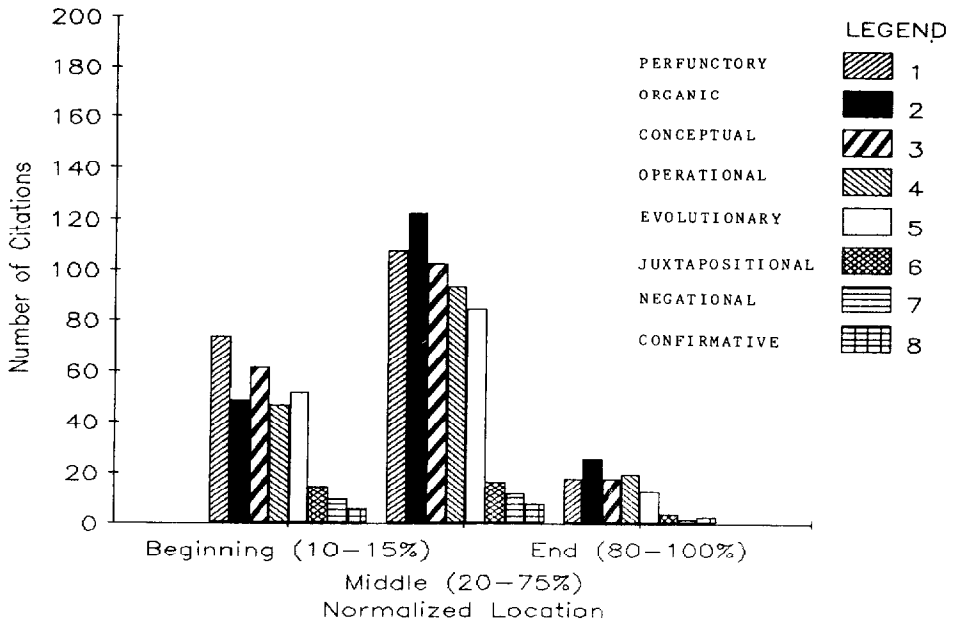


FIG. 4. Citation types over normalized location.

the question of citation location parameters as a new bibliometric element to be used in conjunction with raw citation counts for sensitive, evaluative use of citation measures. The location parameter of a citation is a quantitative variable with potential for qualitative research in citation analysis. Very little research has been done in this area [11].

The major citation type used was the perfunctory although it exhibited a very low utility level. The group of research subjects otherwise seemed to favor equally fundamental and technique information type documents, (organic and operational types). The research subjects showed a marked tendency to group citations in pairs that, according to Moravcsik and Murugesan were supposed to be dichotomous. The research subjects apparently did not perceive any contradiction in their reported use of the information.

The Moravcsik and Murugesan citation behavior model could not accommodate the nature of the information use reported and did not reflect the research subjects' perception of their use of information. However, the citation behavior of scientists in other fields or non-elite scientists in the same field, may show a different information use pattern.

TABLE 4. Utility level per citation type.

| Citation type | 4 | 3 | 2 | 1 | Mean |
|---|---|---|---|---|---|
| Perfunctory | 0.06 | 0.08 | 0.26 | 0.61 | 1.61 |
| Organic | 0.43 | 0.28 | 0.18 | 0.11 | 3.03 |
| Conceptual | 0.37 | 0.32 | 0.23 | 0.08 | 2.98 |
| Operational | 0.46 | 0.32 | 0.15 | 0.08 | 3.18 |
| Evolutionary | 0.35 | 0.32 | 0.19 | 0.15 | 2.89 |
| Juxtapositional | 0.29 | 0.17 | 0.25 | 0.29 | 2.46 |
| Confirmative | 0.39 | 0.11 | 0.22 | 0.28 | 2.61 |
| Negational | 0.20 | — | 0.20 | 0.60 | 1.80 |

1 = Peripheral
2 = Moderate
3 = Heavy
4 = Essential

Bibliometrics and particularly citation analysis hold to the notion that information is represented by discrete, measurable, visible packages of roughly comparable value, (i.e., citations) [12]. Moreover, citations also constitute reflections of intellectual influence. These are central propositions to the field. However, these propositions have recently been the object of serious controversy.

The proposition concerning the visibility of intellectual influence as reflected by citations has recently been contested by MacRoberts and MacRoberts [13]. Their study points out the inability of citations to capture intellectual influences embedded in inarticulated knowledge. In other words, citations can only reflect influences that are specifically acknowledged as such, by the authors of a document; citations do not reflect influences that belong to the "intellectual baggage", accumulated by years of intellectual pursuit. In other words, citations do not capture information or knowledge that has been integrated into the mental constructs of a writer to the extent that it has become inarticulated, or tacit.

Similarly, the findings presented here contradict the notions of discreteness and equality of value of citations. The empirically derived data suggest that a document may contain many items of information that may be cited for a number of reasons. It is therefore impractical to use the citation as a unit of measurement for the study of information transfer and information use. These results agree with Brookes' [7] assessment of the complexities of citers' motivations. His empirical study of 26 academics reported that the motivations of citers cannot be treated "as an undifferentiated whole". The notions of discreteness and equality of value of citations need to be revised, for adequate models of citation behavior to evolve. Indeed, studies focusing on elements of information cited, as opposed to full citations, are needed; particularly to develop models that reflect more adequately the patterns of articulated information use of a scientific community.

## Acknowledgments

Thanks are due to N. Lind and E. Burke for helpful comments and discussions. A preliminary version of this article was presented at the Ninth Annual Meeting of the Society for the Social Studies of Science, Ghent, Belgium, October 1984.

## References

1. Larsen, J. K. "Knowledge Utilization: Current Issues," In: R. F. Rich, Ed. *The Knowledge Cycle*. Beverly Hills: Sage; 1981.
2. Cole, J. R. and Cole, S. *Social Stratification in Science*. Chicago: University of Chicago Press; 1973.
3. Collins, H. M. "Tacit Knowledge and Scientific Networks," In: B. Barnes and D. Edge, Eds. *Science in Context: Readings in the Sociology of Knowledge*. Cambridge: M. I. T. Press; 1982.
4. Small, H. "Citation Context Analysis," In: B. J. Dervin and M. J. Voigt, Eds. *Progress in Communication Sciences. Volume III*. Norwood, N. J.: Ablex; 1982.
5. Brookes, T. A. "Private Acts and Public Objects: An Investigation of Citer Motivations." *Journal of the American Society for Information Science*. 36(4):223–229; 1985.
6. Prabha, C. G. "Some Aspects of Citation Behaviour: A Pilot Study in Business Administration." *Journal of the Society for Information Science*. 34:202–206; 1983.
7. Brookes, T. A. "Evidence of Complex Citer Motivationas." *Journal of the American Society for Information Science*. 37:34–36; 1986.
8. Moravcsik, M. J. and Murugesan, P. "Some Results on the Function and Quality of Citations." *Social Studies of Science*. 5:86–92; 1975.
9. Cano, V. "Communication Patterns of a Scientific Elite." Unpublished Mimeograph. London, Ontario: School of Library and Information Science, University of Western Ontario; 1983.
10. Voos, H. and Dagaev, K. S. "Are all citations equal? or did we op. cit. your idem?" *Journal of Academic Librarianship*. 1:19–21; 1976.
11. Bertram, S. "Citation Counts." In: A. B. Piternick, Ed. Proceedings of the *American Society for Information Science*. Western Canada Chapter. Vancouver, B. C.: University of British Columbia; 1972.
12. Collins, H. M. "The TEA Set: Tacit Knowledge and Scientific Networks." *Science Studies*. 4: 165–186; 1974.
13. MacRoberts, M. H. and MacRoberts, B. R. Quantitative Measures of Communication in Science: A Study of the Formal Level." *Social Studies of Science*. 6(1):151–172; 1986.