The J-shaped distribution of citedness

Jeppe Nicolaisen
Department of Information Studies, Royal School of Library and Information Science, Copenhagen, Denmark

Keywords *Statistics, Libraries, Information*

Received 28 August 2001
Revised 28 January 2002
Accepted 14 March 2002

Abstract *A new approach for investigating the correlation between research quality and citation counts is presented and applied to a case study of the relationship between peer evaluations reflected in scholarly book reviews and the citation frequencies of reviewed books. Results of the study designate a J-shaped distribution between the considered variables, presumably caused by a skewed allocation of negative citations. The paper concludes with suggestions for further research.*

Introduction

We know that scientists differ significantly in productivity and scientific influence. Lotka (1926) observed that the productivity distribution of scientists is extremely skewed, and formulated his well-known inverse square law, stating that the number of scientists producing n papers is roughly proportional to $1/n^2$. Citedness has not been subject to the same inspection as scientific productivity, but available studies are nevertheless indicating that distributions of citations to scientific papers may be as skewed as productivity distributions (Seglen, 1992). Studies of the predictive validity of citation analysis have tested the hypothesis that works with high citation frequencies are valuated more positively than others with lower citation frequencies, and the aim of these previous examinations has consequently been to demonstrate the existence of a linear relationship between research quality and citation counts. But the majority of studies focussing on the quality of documents have only been able to report low to moderate findings of linear correlation, which leads Bornstein (1991) to the conclusion that:

> If a relationship between citation frequency and research quality does exists, this relationship is not likely to be linear. The relationship between research quality and citation frequency probably takes the form of a J-shaped curve, with exceedingly bad research cited more frequently than mediocre research (e.g. as an example of an idea or line of research that turned out to be a blind alley, or as an example of what not to do in a particular area).

Cronin and Overfelt (1994) have actually demonstrated the existence of a J-shaped relationship between researchers' salaries and the number of times

The author gratefully acknowledges the help and support of Professor Peter Ingwersen who supervised an earlier version of this paper. Also many thanks to all of the participants from the Nordic PhD-course in Bibliometrics 2001 for their productive comments on earlier drafts. Finally the author would like to thank three anonymous referees for their constructive critique and suggestions.

Journal of Documentation
Vol. 58 No. 4, 2002, pp. 383-395.

DOI 10.1108/00220410210431118

researchers are cited, and the authors also noted the inappropriateness of assuming linearity. This finding obviously yields some support to Bornstein's hypothesis although it is open to question if salary is a reasonable indicator of quality. We consequently sense the need for further empirical testing.

Whether citation frequency is a useful measure of research quality depends, of course, on how the elusive concept of quality is defined. Based on the philosophical view that there is no absolute truth, and that truth is socially determined (Kuhn, 1962), we agree with Cole and Cole (1973) that current citations are not a measure of the absolute quality of work, but an adequate measure of the quality of work socially defined.

The aim of this paper is not to go further into the discussion of what constitutes research quality, but to present a new technique for investigating the relationship between stated research quality and citation counts. The new research design is based on calculations of median instead of means data and application of the chi-square test instead of linear regression analysis, consequently avoiding the methodological problems of earlier studies. We will employ this technique in a case study of the relationship between peer evaluations reflected in scholarly book reviews and the citation frequencies of reviewed books, and attempt to verify Bornstein's theory of a J-shaped distribution of citedness.

Previous research

The first recorded citation analysis, which applied mathematical and statistical methods to the study of reference lists in scientific documents, is Gross and Gross (1927). The husband and wife team suggested the use of citation analysis as a tool for developing collections of journals and designed and carried out an empirical study in which they employed citation counts as a quality measure. By tabulating the references from the 1926 volume of *Journal of the American Chemical Society*, which they believed to be the most representative journal of American chemistry, they found that 247 different periodicals or journals had been cited. Trends were subsequently traced from these tabulations and employed by the Grosses in the process of assembling a collection of journals for a simulated chemical library. Gross (1927) commented on the applied procedure in a later article:

> Construction of the method of investigation [citation counts] here employed will show that we are not concerned merely with the quantity of work published [. . .] but, that in reality we are concerned only with the good work, the work which has survived and which has proved of value to the investigators who followed.

The best and most valuable works are accordingly those which are repeatedly cited by members of the scientific community, and Gross (1927) was therefore the first to equate scientific quality with the number of citations that a work receives. He did not, however, validate his statement empirically and the Grosses made no attempt to validate the results of their findings. A very important question was consequently left open: is citation analysis really a

valid measure of scientific quality? According to Garfield *et al.* (1978), standards against which the validity of citation analysis can be measured do not exist, and they conclude that all we can do is compare the results of different methodologies, and attempt to find significant correlations between them. Various researchers have previously conducted such investigations, but the majority of them have focussed on either the quality of scientists (e.g. Garfield, 1970; Cole and Cole, 1973), laboratories or departments (e.g. Bayer and Folger, 1966; Narin, 1976); Oppenheim (1995, 1997) or journals (e.g. Garfield, 1972). Only a limited number of researchers have focussed on the quality of single documents and, unfortunately, most of these studies are suffering from a number of methodological flaws:

- documents are drawn from opposite extremes of citation rates and papers in the middle of the distributions are excluded;
- a restricted number of experts are judging the quality of the documents under study;
- binary evaluations;
- mathematical and statistical errors;
- the co-variation between stated research quality and citation counts are measured by linear regression analysis.

Some researchers have only examined documents drawn from opposite extremes of citation rates. Virgo (1977), for example, compared a group of experts' judgements of a number of articles with the frequency with which the same articles were cited. In her experiment she asked the experts to select the most important article in sets of two containing one paper with a high citation frequency and one with a low frequency. She found a low (0.37), but statistically significant, correlation between the experts' evaluations and the citation counts, but admitted that the experiment was rather extreme compared to the normal practice of selecting scientific literature and suggested that future investigations should compare experts' evaluations with results of citation analyses of articles with more varied citation frequencies.

Other researchers have asked a restricted number of experts to valuate the documents under study, and the reliability of these studies may consequently be low. Nicolaisen (2000), for example, employed the opinion of only a single authority, which possibly explains to some degree why the results of his study proved only moderate correlation between peer evaluation and citation counts.

Most studies have operated with binary evaluations exclusively, and the experts have consequently been forced to classify a large proportion of documents to categories where they really do not belong, possibly causing a weaker correlation between the measured variables.

Mathematical and statistical errors are also apparent in some studies. Lawani and Bayer (1983), for example, pooled categories in their chi-square test with expected values below five. The assumed size of the test (assumed

probability of rejecting the null hypothesis) is consequently less than the actual size of the test, and the probability of rejecting the null hypothesis (that there exists no relationship between the measured variables) when it is true is therefore greater than it should be.

Many previous studies have made use of linear regression analysis as a measure of the co-variation between stated research quality and citation counts. Schubert *et al.* (1984), for example, used a similar test when they measured the relationship between the evaluative content of scholarly book reviews and the citation frequencies of the reviewed books. There are, however, various reasons why linear regression analysis is an inappropriate measure in studies of the predictive validity of citation analysis. To begin with, the effects of the independent variables in a linear regression analysis are measured by the differences between means of Y, calculated for the different values of the independent variables. It is, however, a well-known statistical fact, that outliers inflate the mean much more than other measures of central tendency, and the authors of a widely praised statistical textbook (Kvanli *et al.*, 1996) therefore conclude:

> When dealing with data that are likely to contain outliers . . . the median usually is preferred to the mean as a measure of central tendency, since the median provides a more "typical" or "representative" value for these situations.

The regression plot of Schubert *et al.* (1984) clearly exemplifies that citation counts contain outliers and the results of their study may therefore be invalidated by the applied statistical test. Chubin and Moitra (1975) have also observed that citation data contain outliers and they consequently computed the median instead of the mean citation frequency for documents in their comparative study of reference types and citation frequency.

Second, the dependent mean of Y in a linear regression analysis is assumed to be a linear function of the values of the independent variables. But there exists no strong evidence for a linear relation between the stated quality of scientific documents and their citation frequencies. Baldi (1998) actually justified his calculations of the natural logarithm of citations instead of the raw numbers when measuring the quality of potential cited articles, partly by declaring that he did not expect the effect of quality Gas indicated by citation counts Gto be linear. Studies focussing on documents typically show statistically significant results of linearity in the range of 0.10 to 0.40 (e.g. Virgo, 1977; Gottfredson, 1978; Wolfgang *et al.*, 1978; Schubert *et al.*, 1984; Nicolaisen, 2000).

Finally, as judgements reflected by authors usually assume an ordinal scale of ratings, linear regression analysis is scarcely ever the appropriate statistical test since it is a method for specifying the relationship between interval variables (Frankfort-Nachmias and Nachmias, 1996).

Research design

The appropriate procedure for testing the relationship between research quality and citation counts is generally the chi-square test. This test is not very

often used in informetrics (Egghe and Rousseau, 1990), but should be preferred because it can be applied to ordinal data and because it can be employed in tests of non-linear relations. The following investigation demonstrates the application of a chi-square test to a case study of the relationship between peer judgements reflected in scholarly book reviews and the citation frequencies of the reviewed books.

Collection of data

Research on the relationship between research quality and citation counts requires three different collections of data[1]:

(1) A collection of test documents.

(2) A collection of evaluations concerning the documents under study.

(3) A collection of citations to the same documents.

Despite the fact that the journal article has been the preferred forum for scientific communication since the seventeenth century, the monograph is still frequently used in the social sciences and humanities. In the present study we have therefore confined the collection of documents to sociological monographs, which have been reviewed in *Contemporary Sociology* (*CS*) G the most highly regarded book-reviewing journal in sociology (Quist, 1986). This journal is published bi-monthly and contains reviews of newly published scientific works from sociology and related disciplines. It is indexed in *Sociological Abstracts* (*SA*), a database, which provide simple evaluative ratings to book review records. These ratings are specified by the editors of *SA* who read the book reviews and assess the reviewers' general opinions of the books under review on a five point rating scale:

(1) Very favorable.

(2) Favorable.

(3) Neutral.

(4) Unfavorable.

(5) Very unfavorable.

We have employed these evaluative ratings of sociological monographs to the collection of evaluations and collected a proportionate stratified random sample from the five rating categories in *SA*. The sample consists of 5 percent of the book reviews published in *CS* between 1985 and 1994 or at least 40 reviews from each rating category (see Table I). We generated the collection of citations to the 420 monographs in the sample by carefully conducted on-line searches in *Social SciSearch* (*SSCI*). The monographs were looked up as cited reference strings and subsequently checked for alternative listings. Citations from book reviews were omitted from the collection. The reliance on *SSCI*-based citation counts makes the collection of citations a fractionised sample, which may limit the generalizability of the results, since many citations to sociological

monographs probably are located in other monographs beyond the data collection pale (Cronin *et al.*, 1997).

Data handling and analysis

SPSS (Statistical Package for the Social Sciences) and Microsoft Excel were used as tools for the handling and analysis of data sets. The mean and median rates of citations to the sampled monographs in the five rating categories were calculated for later comparison using a five-year citation window, and the strengths of the relationship between the variables was tested by a chi-square test in which citations were pooled in quartiles of overall citation frequency.

Results

The mean and median frequencies of citations to monographs in the five rating categories are calculated in Table II and illustrated in Figures 1 and 2. The two figures signal the quite disparate results obtained by the applied measures of central tendency and simultaneously the inappropriateness of assuming linearity. While the specified line between the mean marks describes a somehow twisted S, the median marks describe a proper J-shaped curve.

Table III illustrates the quartiles in the citation data. The monographs in the five rating categories are grouped in quartiles of overall citation frequency as illustrated in Figure 3. Almost half of the monographs in the very favorable category are located in the fourth quartile and only one fifth of the monographs in the first quartile. The monographs in the unfavorable category have the lowest median citation frequency and are also mainly located in the first quartile with only 6.8 percent of them to be found in the fourth quartile.

Table I.
Sample size

Rating	Number of reviews in *CS* (1985-1994)	5 percent	Sample size
Very favorable	275	13.75	40
Favorable	4,907	245.35	246
Neutral	996	49.80	50
Unfavorable	879	43.95	44
Very unfavorable	70	3.50	40
Total	7,127	356.35	420

Table II.
Computation of the mean and median citation frequencies

Rating	Monographs	Mean	Median
Very favorable	40	17.53	17.50
Favorable	246	19.27	8.00
Neutral	50	15.42	6.00
Unfavorable	44	18.48	4.50
Very unfavorable	40	15.78	6.00
Total	420	17.18	8.00

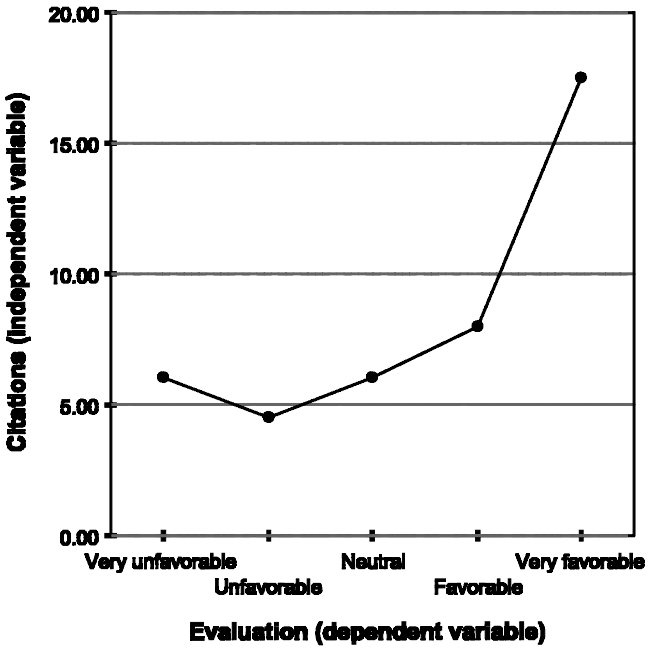

Figure 1.
The median citation frequency

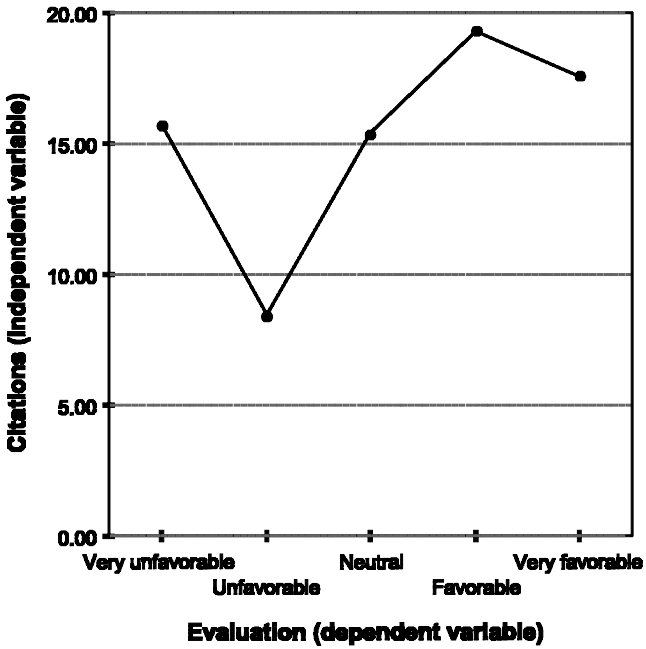

Figure 2.
The mean citation frequency

To test whether the relationship between the variables in Figure 3 is statistically significant, the following statistical hypotheses are tested:

H_O. The distribution of the independent variable (citations) is equally scattered in the categories of the dependent variable (evaluation).

H_A. The distribution of the independent variable (citations) is not equally scattered in the categories of the dependent variable (evaluation).

The results of the chi-square test are presented in Table IV. The chi-square value of the test is 24.4430 and H_O is therefore rejected because $\chi^2 > \chi^2_{\alpha,df} \Longrightarrow \chi^2_{\ 0.05,12} \Longrightarrow 21.0261$. The chi-square test also renders the existence of a highly significant relationship ($p < 0.0185$) between peer evaluations reflected in scholarly book reviews of sociological monographs and the citation frequencies of the reviewed books. The test, however, does not

Quartiles	Number of citations
1	3.00
2	8.00
3	18.75
4	189.00

Table III.
Quartiles in the citation data

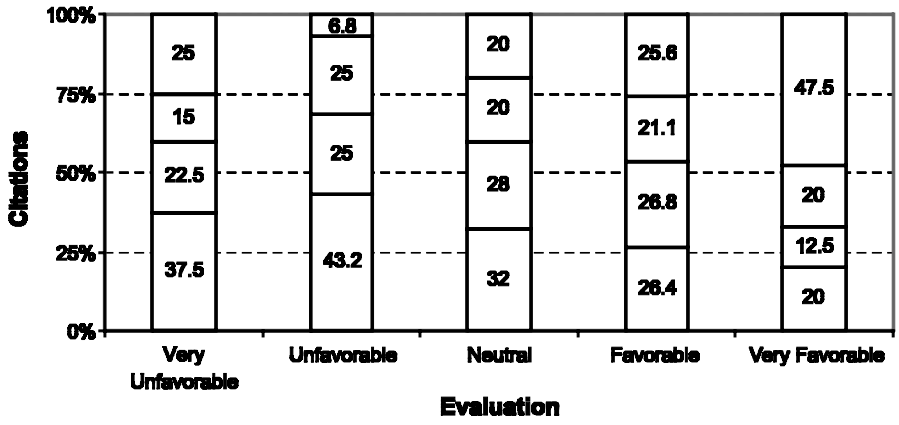

Figure 3.
Monographs grouped in quartiles of overall citation frequencies

	Value	df	Asymp. sig. (two-sided)
Pearson chi-square	24,443[a]	12	0.018
Likelihood ratio	25,447	12	0.013
Linear-by-linear association	10,659	1	0.001
Number of valid cases	420		

Table IV.
Chi-square test

Notes: [a] 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.29

indicate the kind of relationship between the variables in the sample. This has to be interpreted from the median data presented in Figure 1 and Table II. Visual examination of the data strongly indicates that the relationship in fact is J-shaped.

Discussion

Researchers have investigated motives for references or the citing behaviour of scholars by content analyses (e.g. Moravsik and Murugesan, 1975; Chubin and Moitra, 1975), by questionnaires (e.g. Brooks, 1985; Cano, 1989; Shadish *et al.*, 1995; Case and Higgins, 2000), or by interviews (e.g. White and Wang, 1997; Kim, 2000). There exists, unfortunately, no large-scale investigations of when the critical attitude of scholars has lead to heavy criticism of particular works and indirectly to the allocation of numerous negative citations. Garfield (1978a, b), however, has investigated the types of citations received by a highly debated article by Arthur Jensen in the 1969 volume of *Harvard Educational Review*. Jensen's article was according to Garfield (1978b) heavily cited because it had been seriously criticised. Of the 60 papers in Garfield's sample, 29 cited Jensen's article negatively. Eight cited the article as an example of a controversy. Eight more used the article as a background reference, while only 15 of the articles cited Jensen in agreement with his positions, and seven of them only on minor points. In a later study, Garfield and Welljams-Dorof (1990) investigated the impact of fraudulent research on scientific literature by focussing on 20 publications from the Steven E. Breuning case. The authors retrieved copies of 65 citing articles and identified in each text where Breuning's work was cited and how it was referred. Findings of their content analysis indicated that less than 10 percent of the citations were positive of nature. Regarding the Jensen work, Garfield (1978b) speculated that scientists probably felt they could not disregard it, but concluded that specific examples of high impact papers, which have a large percentage of critical or negative citations, are rare. This corresponds to the thoughts of Meadows (1974) who found that surprisingly enough, despite its acceptance of the need for organised scepticism, the scientific community does not normally go out of its way to refute incorrect results. According to Meadows (1974):

> This lack of activity stems partly from the recognition that, although a few scientists have made their names by a detailed criticism of existing literature, most kudos goes to those who add something positive to scientific knowledge; but partly it is due to an implicit recognition that refutation is often unnecessary. If incorrect results stand in the way of the further development of a subject, or if they contradict work in which someone else has a vested interest, then it may become necessary to launch a frontal attack. Otherwise, it generally takes less time and energy to bypass erroneous material, and simply allow it to fade into obscurity.

Bornstein (1991) accounts for two possible causes of the J-relationship (poor research is cited more frequently than mediocre research as examples of ideas or lines of research that turned out to be blind alleys, or as exemplars of what not to do in particular areas), but fails to mention a third and imaginably more

accurate explanation. We believe that a J-shaped relationship between research quality and citation counts indicates that erroneous research is criticised and referred to negatively by some scholars in their publications and as a consequence is being more frequently cited than mediocre or ordinary research, which is mostly left unnoticed. We consequently perceive the predominant epistemological attitude in a discipline or a field as contributing strongly to the shape of relationship between research quality and citation frequency. It is easy to imagine that if researchers from one field maintain a strong positivistic attitude in the process of research, then they would primarily be citing verifiable results of others, as the positivistic epistemology is driven by the principle of verification. If the researchers instead were guided rigorously by the Popperian falsification philosophy, then they would cite both correct and erroneous results, as this philosophy indebts the researchers to criticise erroneous or misleading research whenever they come a cross such work. This corresponds to Merton's (1942) norm of organized scepticism, which requires that all science, no matter how eminent the scientist, should be judged critically. One may therefore rightfully question if we also would find a J-shaped relationship in the natural sciences, which traditionally are more positivistic oriented than the social and humanistic sciences.

The J-shaped relationship found in the present study indicates that sociological research is neither driven rigorously by the principle of verification nor by the principle of falsification, but instead that the individual authors choose if they will meet the erroneous work of others with critique. It further suggests that sociological researchers only maintain critical attitudes in special cases G when erroneous research cannot be disregarded G as Meadows precisely explained. This suggestion is of course empirically testable. By content analysing an appropriate fraction of the more than 7,000 citing papers in the original collection of citing documents it would be possible to infer whether the negative citations are distributed equally in the five rating categories or if they, as we presume, primarily are located in the class of very unfavorable reviewed books. What is needed, however, is author cooperation or professional assistance by experienced sociologists since some citations possibly are hedged and consequently hard to classify by non-specialists (MacRoberts and MacRoberts, 1984).

Conclusion

Findings of the investigation of the correlation between peer reviews reflected in scholarly book reviews and the citation frequencies of the reviewed books designate a *J*-shaped association. The results of the present study therefore confirm Bornstein's (1991) hypothesis of a J-shaped relationship between research quality and citation counts. The case study is, however, restricted to a sample of sociological monographs, which limits generalizability, since virtually all large-scale citation analyses of note are based on the journal literature, and especially on the literature of the natural sciences. In order to

investigate whether the J-shaped relationship between research quality and citation frequency is universal, and perhaps even a bibliometric law, future studies should therefore seek to explore other disciplines and other types of scientific communication.

Previous investigations have mostly reported low to moderate linear correlations between research quality and citation counts when measured by linear regression analysis. The diverse, J-shaped relationship found in the present study is established by means of a new technique, which is based on calculations of median data and application of the chi-square test. There are three main reasons why this technique is superior to the preceding methods based on means data and application of linear regression analyses:

(1) Citation data contain outliers, which inflate calculations of the mean citation frequency much more than the median frequency.

(2) The relationship between research quality and citation counts is not linear.

(3) Judgements reflected by experts regarding research quality usually assume an ordinal scale of ratings.

Future investigations should therefore compute the median instead of the mean citation frequency and compare the median citation frequency of documents with the evaluations of experts. Prospective studies should moreover make use of the chi-square test when measuring the strength of co-variation between the considered variables because of the test's applicability to ordinal data and because it can be applied to investigations of non-linear relationships.

Note

1. Data are on file. Inquiries may be addressed to the authors.

References

Baldi, S. (1998), "Normative versus social constructivist processes in the allocation of citations: a network-analytic model", *American Sociological Review*, Vol. 63, pp. 829-46.

Bayer, A.E. and Folger, J. (1966), "Some correlates of a citation measure of productivity in science", *Sociology of Education*, Vol. 39 No. 4, pp. 381-90.

Bornstein, R.F. (1991), "The predictive validity of peer review: a neglected issue", *Behavioral and Brain Sciences*, Vol. 14 No. 1, pp. 138-9.

Brooks, T.A. (1985), "Private acts and public objects: an investigation of citer motivations", *Journal of the American Society for Information Science*, Vol. 36 No. 4, pp. 223-9.

Cano, V. (1989), "Citation behavior: classification, utility, and location", *Journal of the American Society for Information Science*, Vol. 40 No. 4, pp. 284-90.

Case, D.O. and Higgins, G.M. (2000), "How can we investigate citation behavior? A study of reasons for citing literature in communication", *Journal of the American Society for Information Science*, Vol. 51 No. 7, pp. 635-45.

Chubin, D.E. and Moitra, S.D. (1975), "Content analysis of references: adjunct or alternative to citation counting?", *Social Studies of Science*, Vol. 5, pp. 423-41.

Cole, J.R. and Cole, S. (1973), *Social Stratification in Science*, University of Chicago Press, Chicago, IL.

Cronin, B. and Overfelt, K. (1994), "Citation-based auditing of academic performance", *Journal of the American Society for Information Science*, Vol. 45 No. 2, pp. 61-72.

Cronin, B., Snyder, H. and Atkins, H. (1997), "Comparative rankings of authors in monographic and journal literature: a study of sociology", *Journal of Documentation*, Vol. 53 No. 3, pp. 263-73.

Egghe, L. and Rousseau, R. (1990), *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*, Elsevier, Amsterdam.

Frankfort-Nachmias, C. and Nachmias, D. (1996), *Research Methods in the Social Sciences*, 5th ed., Arnold, London.

Garfield, E. (1970), "Citation indexing for studying science", *Nature*, Vol. 227, pp. 669-71.

Garfield, E. (1972), "Citation analysis as a tool in journal evaluation", *Science*, Vol. 178, pp. 471-9.

Garfield, E. (1978a), "The 100 articles most cited by social scientists, 1969-1977", *Current Contents*, Vol. 32, 7 August, pp. 5-14.

Garfield, E. (1978b), "High impact science and the case of Arthur Jensen", *Current Contents*, Vol. 41, 9 October, pp. 652-62.

Garfield, E. and Welljams-Dorof, A. (1990), "The impact of fraudulent research on scientific literature: the Steven E. Breuning case", *Journal of the American Medical Association*, Vol. 263 No. 10, pp. 1424-6.

Garfield, E., Malin, M.V. and Small, H. (1978), "Citation data as science indicators", in Elkana, Y. *et al.* (Eds), *Toward a Metric of Science: The Advent of Science Indicators*, J. Wiley, New York, NY, pp. 179-207.

Gottfredson, S.D. (1978), "Evaluating psychological research reports: dimensions, reliability, and correlates of quality judgements", *American Psychologist*, Vol. 33, pp. 920-39.

Gross, P.L.K. (1927), "Fundamental science and war", *Science*, Vol. 66, 20 December, pp. 640-5.

Gross, P.L.K. and Gross, E.M. (1927), "College libraries and chemical education", *Science*, Vol. 66, 27 October, pp. 385-9.

Kim, H.J. (2000), "Motivations for hyperlinking in scholarly electronic articles: a qualitative study", *Journal of the American Society for Information Science*, Vol. 51 No. 10, pp. 887-99.

Kuhn, T. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, IL.

Kvanli, A.H., Guynes, C.S. and Pavur, R.J. (1996), *Introduction to Business Statistics: A Computer Integrated Approach*, 5th ed., West Publishing, St Paul, MN.

Lawani, S.M. and Bayer, A.E. (1983), "Validity of citation criteria for assessing the influence of scientific publications: new evidence with peer assessment", *Journal of the American Society for Information Science*, Vol. 34 No. 1, pp. 59-66.

Lotka, A.J. (1926), "The frequency distribution of scientific productivity", *Journal of the Washington Academy of Science*, Vol. 16, pp. 317-23.

MacRoberts, M.H. and MacRoberts, B.R. (1984), "The negational reference: or the art of dissembling", *Social Studies of Science*, Vol. 14, pp. 91-4.

Meadows, A.J. (1974), *Communication in Science*, Butterworth, London.

Merton, R.K. (1942), "Science and technology in a democratic order", *Journal of Legal and Political Sociology*, Vol. 1, pp. 115-26.

Moravcsik, M.J. and Murugesan, P. (1975), "Some results on the function and quality of citations", *Social Studies of Science*, Vol. 5, pp. 86-92.

Narin, F. (1976), *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*, Computer Horizons, Cherry Hill, NJ.

Nicolaisen, J. (2000), "Combination of citation analysis and expert judgement: a methodology for research evaluation", Internal report, Royal School of Library and Information Science, Copenhagen (in Danish).

Oppenheim, C. (1995), "The correlation between citation counts and the 1992 research assessment exercise ratings for British library and information science university departments", *Journal of Documentation*, Vol. 51 No. 1, pp. 18-27.

Oppenheim, C. (1997), "The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology", *Journal of Documentation*, Vol. 53 No. 5, pp. 477-87.

Quist, S. (1986), "Book reviewing in sociology", *Reference Librarian*, Vol. 15, pp. 75-85.

Schubert, A. *et al.* (1984), "Quantitative analysis of a visible tip of the peer review iceberg: book reviews in chemistry", *Scientometrics*, Vol. 6 No. 6, pp. 433-43.

Seglen, P.O. (1992), "The skewness of science", *Journal of the American Society for Information Science*, Vol. 43 No. 9, pp. 628-38.

Shadish, W.R. *et al.* (1995), "Author judgement about works they cite: three studies from psychology journals", *Social Studies of Science*, Vol. 25, pp. 477-98.

Virgo, J.A. (1977), "A statistical procedure for evaluating the importance of scientific papers", *The Library Quarterly*, Vol. 47 No. 4, pp. 415-30.

White, M.D. and Wang, P. (1997), "A qualitative study of citing behavior: contributions, criteria, and metalevel documentation concerns", *Library Quarterly*, Vol. 67 No. 2, pp. 122-54.

Wolfgang, M.E., Figlio, R.M. and Thornberry, T.P. (1978), *Evaluating Criminology*, Elsevier, New York, NY.