Notes and Letters, *Social Studies of Science*, Volume 7 (1977), 257-67

Citation Analysis: Queries and Caveats

Alan L. Porter

Study of citation patterns has become a leading analytical tool in the sociology of science, and offers an enticing way to evaluate the research productivity of individual scientists. However, there remain some basic conceptual and methodological concerns. Debate touches on the validity of citation counts as an indicator of 'scientific quality' (of papers or authors), of recognition, of scientific diffusion, and of utilization (who cites whom).[1] To Menard, citations are the only enduring record that a paper is used,[2] and Cole and Cole take citations to be a roughly valid indicator of influence.[3] But Kaplan wonders what citations mean, in the absence of knowledge about actual norms and practices.[4] Plausible reasons for citation include selectively supporting one's contentions (i.e. 'persuasion'), indicating knowledge of the literature, providing clues about content to citation index scanners -- as well as acknowledging genuine communication.[5] Citation does not actually prove a reference has been read (or even seen),[6] yet journal citations appear to be the best practical indicator of the worth of research.[7]

Citations may be used to evaluate papers (either individually or in groups — e.g. experimental vs. theoretical physics; or short notes vs. full reports); or, similarly, to evaluate journals; or scientists. This paper concerns citation analysis of scientist-authors (both individuals and groups), and discusses the methodological problems involved.

The problems to be faced depend on the uses to which the measures are to be put. The problems and opportunities of a supervisor evaluating the citation profile of a young scientist (an arguable practice) are poles apart from those of a sociologist studying groups of scientists in an international specialty, or of a historian studying its development. However, there are some generalizable concerns: these include the possibility that radical work will go unrecognized (and uncited) by contemporaries; that outstanding ideas become widely accepted and not formally cited (e.g. the Mossbauer effect);[8] and that citation counts do not

Author's address: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.

distinguish critical commentaries.[9] Other methodological considerations include the appropriate 'weighting' of individual citations; ways to distinguish 'quantity' and 'quality' measurements; and techniques of standardization, to allow comparison of citation patterns across fields, and over time.[10]

This study focuses on three concerns: (1) citations of collaborative research; (2) self-citations and citations among close associates ('collaborators'); and (3) characteristics of citation data. These concerns are particularly important when comparing scientific authors. I am aware of only one quantitative study that substantively addresses the first two points,[11] and relatively little attention in the literature to the implications of the third.

AN EMPIRICAL STUDY

I have compiled citation histories for a sample of 128 psychologists who received their PhDs in the 1963-64 academic year, and who responded to a questionnaire on their dissertations, professional careers, and publications for the years 1964-72. Of the 128, 23 had no publications, and 36 were not explicit on co-authorship status. This leaves 69 who provided the requisite information. The survey covered a random sample of all American psychologists gaining doctorates in 1963-64, but respondents published much more heavily than non-respondents.[12]

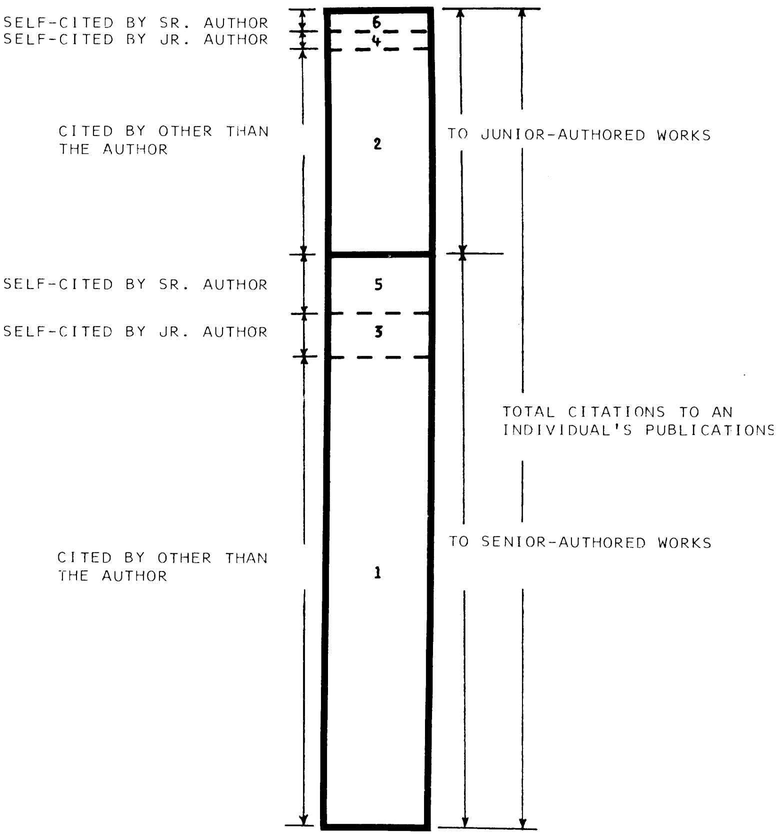
Figure 1 schematically illustrates possible categories of citation. The ideal tally for most purposes would be (1 + 2): citations by other(s) to works for which the individual is senior (sole or otherwise) or junior author.[13] But the easiest is (1 + 3 + 5): citations by anyone to publications of which the individual is senior author (neglecting any junior, co-authored works). And the general practice appears to be a compromise of (1 + 3): citations by others to senior-authored works alone. I wish to assess the implications of excluding citations to junior-authored work (2), while including co-author citations (3).

CITATIONS OF COLLABORATIVE RESEARCH

Citation analysis of collaborative studies is difficult because the *Science Citation Index (SCI)* lists citations only by senior (first) authors. To count citations to junior-authored works requires identification of senior collaborator(s), and a search for specific citations under their names. If this is not done, how serious is the resultant loss of information? The Coles' study of 120 highly productive, high-prestige academic physicists used *Science Abstracts* to trace co-authored publications,[14] and found a correlation of $r = 0.96$[15] between citations to senior-authored works only and total citations. Data on the 69 psychologists coroborate this result, $r = 0.95$.[16]

Cole and Cole also considered whether the practice of listing co-authors alphabetically was a significant factor. If it were, scientists whose surnames begin with early letters would show a higher proportion of senior to total citations, giving an unfair advantage in comparisons based on senior citation counts. But the ratio of senior to total citations for A-M physicists was 67%, and for N-Z, 71% – a slight difference in the opposite direction. For the psychologists

Figure 1. An Average Psychologist's Citation Profile



(excluding self-citations) the corresponding figures were 70% and 58%: but the first figure drops to 62% for the A-G sub-group; and the difference reverses if one counts *publications* – A-M, 65% senior authorship; N-Z, 70%. There is therefore no evidence in either study that alphabetical ordering invalidates senior-author-only citation counts.

Results suggest that senior citations should usually be an adequate substitute for total citations. Of course, this may not be so for individuals or small groups; about 15% of the publishing psychologists were junior authors on 45% or more of their publications. Also the loss of information in this sample was substantial – 32% of the publications (or 33% of the non-self-citations) were junior-authored, and hence not counted; since junior-authorship correlated only 0.56 with senior-authorship, considerable information was at stake. However, the effort required to obtain this missing one-third of the data is, for most purposes, impractical.

To list co-authored publications requires either a questionnaire or a (necessarily incomplete) search of an abstracts index. For a sample of psychologists, the citation search effort would be multiplied about four-fold.

Several potential biases of this attentuated, but practical, strategy can be identified. For one, if younger or less eminent colleagues tend to appear as junior authors more often than their contributions warrant, senior-authored citation counts will further denigrate their contributions to science.[17] In terms of publications, one could speculate that the power law coefficient 'a' in 'Lotka's Law' (which states that the fraction of scientists producing N papers in a field is proportional to $1/N^a$) would be lower if one credited each collaborator's authorship.[18] This leads on to the question of what weighting a co-author deserves. What should be his 'share of the product'? Between the handy weighting of zero (senior-author counts only) and the too-generous weighting of one (all-author counts), come proportional counting (divide the credit by the number of collaborators), or even some form of sliding credit (e.g. first author gets a double share). Before we can sensibly assess these possibilities, we need to know more about the norms governing the allocation of collaborative authorship in the various sciences. Meanwhile, it seems most informative to pursue *multiple approaches*. For instance, in an all-author count, one could also make a senior-author tally, for comparison; or one could use different intermediate weightings and observe how the results change. If counts remain consistent, or shift in an interpretable manner, our confidence in them would increase.

Or consider other studies of scientific productivity. How might the evidence for the dominance of a research élite (as opposed to the 'Ortega Hypothesis', which postulates broad-based contributions to scientific advance) shift if one weighted co-authors' citations less?[19] Again, multiple weighting strategies could prove illuminating. Indeed, it is not clear that giving co-authors more consideration would necessarily broaden the scientific pyramid substantively. Meadows notes that highly productive individuals tend to appear more often as co-authors.[20] Hence, senior-author counts may also have a 'levelling' effect on high producers. Different co-authorship practices can also bias comparisons of productivity. For instance, Meadows and O'Connor found 1.7 times as many authors per paper in observational pulsar research as in the theoretical literature.[21] Hence senior-author citation counts might benefit theorists, while all-author counting would better represent individual experimentalists. The extent of co-authorship, and the number of co-authors, differs over time, and between fields — both tending to increase with time and with a field's 'hardness'.[22] Any study which seeks to assess the productivity of individual scientists by citation analysis is therefore subject to such measurement quirks.

To summarize: cost and convenience together argue strongly in favour of senior-author-only counts. They correlate highly with all-author counts, and studies show no alphabetical anomalies. However, they involve a substantial (one-third) loss of citation information, and this could produce significant bias. Various weighting strategies merit consideration: the choice of an appropriate strategy depends on one's aims and on the nature of the specialty concerned. But the overall recommendation, given the current state-of-the-art, would be to use both all-author and senior-author counts, along with alternative weightings of co-authorship, and carefully compare the results.

Table 1. Detailed Publication and Citation Profiles for a Random 8 of the Sample of 69 Psychologists

	Publications					Citations					
Author	Total	Sole Author	Senior Co-author	Junior Co-author	Unique Senior Collaborators	Total	Number Rechecked	Senior Author Self-Citations	Junior Author Self-Citations	Other Collaborators' Citations	Non-self Citations
A	15	5	5	5	(5)	14	13	0	0	0	17
B	2	0	1	1	(1)	0	0	0	0	0	0
C	4	2	0	2	(2)	14	7	0	0	0	7
D	20	6	6	8	(4)	72	29	3	1	1	33
E	5	1	1	3	(3)	35	17	0	0	0	13
F	49	17	10	22	(13)	189	113	8	8	45	107
G	16	4	4	8	(3)	14	8	0	0	0	8
H	1	0	1	0	(0)	0	0	0	0	0	0

Publications of these 1963 PhDs through 1973. I note subjectivity in such counts. I included cited (but not uncited) technical reports, unpublished presented papers, abstracts, and theses, and reached decisions concerning letter replies, book chapters, edited works, works not yet published, semi-professional articles, discussions, and so on.

Citations derived from *SCI* for 1964-72. Rechecking was performed only for years 1970-72 wherein citations were noted; the breakdown of citations into senior-authored, junior-authored, other collaborator-cited, and non-self-cited pertains to the rechecked citations only. Collaborators' citations, a subset of non-self citations, are the number by persons who have collaborated in some other publication.

SELF-CITATIONS

Turning to self-citation, the overall finding is reassuring; the correlation between total citations with and without self-citations included is 0.99 for the psychologists; for senior-authored works only, 0.98. For some purposes, therefore, one could reasonably include self-citations (16.1% of the total 1920 citations).[23] But again, for specific individuals or small groups, self-citation should be separated; 10% of the psychologists self-cited for 30% or more of their citations.

Table 1 presents further details on eight of the psychologists. It shows that the extent of self-citations is insignificant: so too, with one exception (F), is citation by collaborating colleagues. But F displays an interesting profile – of 189 citations to his work, 81 were made by only 7 people, of whom 4 have been collaborators (45 citations); and 11 of his other 26 'citers' mentioned only a single methodological article.[24] While such detail may occasionally be important in the sociology of science, the overall conclusion is that separate determination of citations by co-authors is not worth the effort.

Table 2 considers citation patterns for sub-groups of the psychologists. Results *suggest* that the closer their work approximates the model of the academic natural science laboratory, the greater both the citation productivity and collaboration in publication, and the lower the rate of self-citation. Meadows observes similar differential rates of self-citation across fields and over time.[25] He suggests that the self-citation proportion may reflect the 'maturity' of a specialty. The number and proportion of non-self-citations might then indicate the maturation of an individual's research, as others begin to notice it.

Table 2. Differences Among Psychologists in Citation Patterns

	N	Non-Self Citations *Mean*, Median	% Senior Citations	% Self Citations
Clinical/Counselling	33	9.6, 1.4	72	22.5
Non-Lab	17	13.4, 5.0	68	15
Lab	19	56.4, 35.0	65	14
Non-Academic	25	9.2, 1.3	78	25
Academic	44	31.4, 14.0	65	14

Classification of psychologists is based upon their questionnaire responses (see note) – *Non-Lab* includes developmental, educational, industrial or engineering, personality, psychometrics and social specialties; *Lab* includes learning, physiological, and experimental. Cell sizes are too small for more detailed analysis. *% Senior citations* calculated from non-self citations; results are similar using total citations. *% Self-citations* calculated from total citations; results are similar using only senior citations. Percentages are based on ratios of summations and hence variance indicators are not appropriate. One-way χ^2 tests for each of the sub-groupings were significant for non-self citations, but not for % senior citations and % citations.

The interpretation of self-citations raises alternatives. Self-citations could be used to note the most relevant earlier work, or work with which one is most familiar; they could then reflect real influences and contributions. But there are also potentially self-serving aspects of self-citation, such as propounding one's own work and/or accumulating citations. Co-author citations offer additional self-serving possibilities – better 'camouflaged' citation accumulation, and 'back-scratching'. But these reasons for self-citation are not qualitatively different from those for any citations. So study of citation norms in general could help us in handling self-citations. Meanwhile, selection of an appropriate strategy should reflect the researcher's aims. Present correlations indicate little difference if self-citations are excluded – but one then loses some 10% of the citation information, and this *could* severely alter certain detailed, individual comparisons. Besides, self-citation rates may themselves provide useful information on such topics as the 'maturity' of a field.

CHARACTERISTICS OF CITATION DATA

Two points should be noted. First, these citation data are extremely positively skewed. For the 69 psychologists, citation skewness was over 2.0 for total, senior, and junior citations. Respective means and medians (for non-self-citations) were: 23.3 and 7.2 for total citations: 15.5 and 4.2 for senior; and 7.8 and 0.5 for junior (and these data exclude the 23 with no publications!). This reflects the fact that most papers are never cited, but a few are cited extensively, particularly new technique papers.[26] The most cited author of 1967 owed this standing to 2,350 citations to a single 1951 publication (incidentally written with 3 co-authors).[27] Thus, statistics based on the assumption of normal distribution should be applied with caution.

Others have noted this statistical property of citation data – whether citations of individuals, articles, or journals. Some suggest the use of median, rather than mean, citations, which reduces the influence of the extremely high counts (the positive skew);[28] others, the use of rank order data (as in Spearman's rho), which does likewise.[29] More general concerns are reflected in the variety of suggested schemes to count and categorize data – dividing the alphabet in half to examine the effect of author ordering; counting citations to a year's output, and weighting them by the age of the cited paper;[30] categorizing citations as 'basic', 'subsidiary', 'additional information', or 'perfunctory';[31] counting the number of papers citing an individual at least once;[32] and so on. This variety implies a variety of underlying theoretical notions. Citations are measurable indicators, logically linked to interesting theoretical variables (e.g. scientific productivity, communication units, or whatever), but the correct functional form of this linkage is unknown. At best, the errors in measurement thus introduced may not be serious: but at worst, the indicator may appear inappropriate – 'construct validity' is lacking.[33]

This leads us to the second point. The presence of errors is a non-trivial factor; initial citations, *SCI* compilation, and search of *SCI* data, are all suspect. For instance, in rechecking citations for Table 1 (for a mere 6 cited authors, for three years), I noted problems of two sorts. First, we introduced errors in the search through miscounting the number of citations of an article, failing to identify self-citations, and counting citations of an article by someone else. Second, we found inconsistencies beyond our control: an author's own list of publications omits articles, or gives incorrect details; wrong page or volume cited; technical reports cited in very different ways; impossible citations (but probably correct authors); and authors changing name (married name). The 'homograph' (duplicate name author in *SCI*)[34] presents a major problem for automated searches – possibly losing 25% of a sample, if the US National Research Council's experience with a file of roughly 500,000 scientists is typical. Computing Table 1 uncovered many discrepancies with the original figures, reflecting the difficulty of the task. Adding errors in initial citations and *SCI* compilation, a random error rate of nearly 20% was probably present in our study. Cole and Cole, in discussing the existence of such errors, conclude that citation counts have value as indicators of the quality of a scientist's work, but that interpreation of fine differences would be unwise.[35] Chubin and Moitra also note errors, both random and systematic – such as underestimation of citations since preprints can only be indexed under 'in press'.[36]

IMPLICATIONS

How does the 'signal' retained vary with the type of citation count performed? If one loses 33% of an individual's citations (co-authored ones), fails to cull 5% co-author self-citations, and adds a substantial random error rate, what does one have? Automated search may surrender 25% of the sample. Further, estimates of the proportion of 'perfunctory' citations in high-energy physics range from about 20 per cent[37] to 40 percent.[38] The value of a citation count is thus a complex function of the type of counting and the intended uses; it cannot be taken at simple face value.

But let us look at the overall citation analysis process, as applied to individual (or groups of) scientists.[39] To illustrate the issues that arise, let us focus on a particular citation study – the research on 'The Ortega Hypothesis' by Cole and Cole.[40] The Coles explicitly acknowledge the premise that the scientific establishment is stratified, and state their research question as to 'whether progress is built on the labor of all "social classes", or is primarily dependent on the work of an "élite".' They operationalize this by asking how many scientists are contributing through their published research, assuming 'that the research cited by scientists in their own papers represents a roughly valid indicator of influence on their work.' They then select samples of scientists, and conduct an empirical investigation.

Their study raises questions that in more or less modified form apply to any citation study.[41] One presumably begins with an interest in some underlying abstractions (some theoretical variables) and the ways they interplay. To allow research to proceed, such a model must be simple enough to limit the number of relevant variables and interactions. Is the notion of a 'stratified scientific establishment' meritorious? Is 'scientific progress' a meaningful concept? Is the model relating productivity contributions to stratification level sufficiently rich (complex) to represent reality? For operationalization, one must decide what to measure, and how to relate those measurements to the variables. Are strata best represented as dichotomous, élite, or other? Do citations fairly represent 'influences'? Do publications appropriately indicate 'productivity'? How would results differ if 'productivity' were measured by technological innovation, and only related publications were to count?

How adequately do the measurements test the hypothesis? Are there systematic biases, making conclusions suspect? For instance, if close colleagues selectivity cite each other's work for reasons other than scientific influence, will not results be biased against the Ortega Hypothesis? Or would conclusions be the same if co-author and senior-author citations were not weighted equally? If there is sufficient random error, true effects may be masked. Might the presence of perfunctory citations and tabulation errors lessen the statistical significance of the results? Definition of the strata population is difficult; and the samples studied may not be representative. How general are the findings? Are influences upon physicists like those upon sociologists?

Because substantial, systematic count differences arise when self-citations and citations to junior-authored works are handled differently, comparison of published citation analyses requires close attention to the definitions and practices used. Shifting citation base rates, and variance in scholarly practices, can make comparative citation analysis (over time, or between specialties) a precarious business. Present findings support the cautious use of hand or automated citation counting of senior-authored papers, without concern for co-author self-citations – provided the researcher carefully considers the implications for his analysis of his choice of measures, of potential biases, and of loss of information.

NOTES

I thank Dr Dael Wolfle, University of Washington and Dr Morton V. Malin, Institute for Scientific Information, for helpful review and advice. The E.W. Hazen Foundation and the Graduate School Research Foundation of the University of Washington graciously supported the research.

1. See, for instance, N. Kaplan, 'The Norms of Citation Behavior: Prolegomena to the Footnote', American Documentation, Vol. 16 (1965), 179-84; J.R. Cole and S. Cole, Social Stratification in Science (Chicago: The University of Chicago Press, 1973); M.J. Moravcsik and P. Murugesan, 'Some Results on the Function and Quality of Citations', Social Studies of Science, Vol. 5 (1975), 86-92.
2. H.W. Menard, Science: Growth and Change (Cambridge, Mass.: Harvard University Press, 1971), 26.
3. J.R. Cole and S. Cole, 'The Ortega Hypothesis', Science, Vol. 178 (27

October 1972), 368-75.

4. Kaplan, op. cit. note 1.

5. See G. Nigel Gilbert, 'Referencing as Persuasion', Social Studies of Science, Vol. 7 (February 1977), 113-22; and S.A. Goudsmit, 'Citation Analysis' (Letter to the Editor), Science, Vol. 183 (11 January 1974), 28.

6. D. Davies, 'Citation Idiosyncracies?' (Letter to the Editor), Nature, Vol. 228 (26 December 1970), 1356.

7. K.E. Clark, America's Psychologists: A Survey of a Growing Profession (Washington: American Psychological Association, 1957); A.E. Bayer and J. Folger, 'Some Correlates of a Citation Measure of Productivity in Science', Sociology of Education, Vol. 39(1966), 381-90.

8. On this point, see the Note by Daniel Sullivan and his colleagues in this issue: 'Co-Citation Analyses of Science: An Evaluation', Social Studies of Science, Vol. 7 (May 1977), 223-40.

9. D.E. Chubin and S.D. Moitra, 'Content Analysis of References: Adjunct or Alternative to Citation Counting?', Social Studies of Science, Vol. 5 (1975), 423-41, found only around 5% partially negative references in a sample of high-energy physics articles.

10. See Cole and Cole, op. cit. note 1.

11. Ibid., Appendix A, 263-65.

12. A.L. Porter and D. Wolfle, 'Utility of the Doctoral Dissertation', American Psychologist, Vol. 30 (1975), 1054-61.

13. M.V. Malin (personal communication) emphasizes the hazard in neglecting co-author citations, despite the additional effort involved.

14. Cole and Cole, op. cit. note 1, Appendix A, 263-65.

15. Spearman's rho = 0.85.

16. Spearman's rho = 0.96.

17. A.J. Meadows, Communication in Science (London: Butterworths, 1974), 196-97, raises several other complications, such as the variation of graduate student authorship position and differences in the authorship credit given to skilled technicians.

18. Meadows, ibid., 192, discusses Lotka's Law: see also D.J. de S. Price, Science Since Babylon (New Haven, Conn.: Yale University Press, 1961), and Paul D. Allison et al., 'Lotka's Law: A Problem in Its Interpretation and Application', Social Studies of Science, Vol. 6 (May 1976), 269-76.

19. Cole and Cole, op. cit. note 3, traced out all authors on cited papers for whom they could obtain information.

20. Meadows, op. cit. note 17, 195-98.

21. A.J. Meadows and J.G. O'Connor, "Bibliographical Statistics as a Guide to Growth Points in Science', Science Studies, Vol. 1 (1971), 95-99. Chubin and Moitra, op. cit. note 9, found 6.9 times as many authors per paper for a sample of experimental as opposed to theoretical high-energy physics articles. See also the data presented elsewhere in this issue by Sullivan and his colleagues: 'The State of a Science . . .', Social Studies of Science, Vol. 7 (May 1977), 167-200.

22. Meadows, op. cit. note 17, 199-206.

23. Self-citations here include citations by co-authors of a given article. Others report self-citation rates ranging from 7% (Chubin and Moitra, op. cit. note 9, 430), to 8-10% (Kaplan, op. cit. note 1), on up to about 20% (Meadows and O'Connor, op. cit. note 21). Meadows (op. cit. note 17, 160-61) suggests 10%

as a general approximation, somewhat lower than the present figure, but quite reasonable given the present inclusion of co-author self-citations.

24. This is reminiscent of Menard's description (op. cit. note 2) of who had cited his work, that many cited his work quite heavily, and that he knew essentially all of his heavy citers; i.e. an 'invisible college' existed.

25. Meadows, op. cit. note 17, 159-61.

26. Anonymous (Editorial), 'More Games with Numbers', Nature, Vol. 228 (21 November 1970), 698-99.

27. E. Garfield, 'Citation Indexing for Studying Science', Nature, Vol. 227 (15 August 1970), 669-71.

28. Chubin and Moitra, op. cit. note 9.

29. Cole and Cole, op. cit. note 1, Appendix A.

30. J.R. Cole and S. Cole, 'Measuring the Quality of Sociological Research: Problems in the Use of the Science Citation Index', American Sociologist, Vol. 6 (1971), 23-29. They weight citations 17-fold if the cited paper appeared twenty or more years before citation.

31. Chubin and Moitra, op. cit. note 9.

32. G.J. Stigler and C. Freidland, 'The Citation Practices of Doctorates in Economics', Journal of Political Economy, Vol. 83 (1975), 477-507.

33. T.D. Cook and D.T. Campbell, 'The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings', in M.D. Dunnette (ed.), Handbook of Industrial and Organizational Research (New York: Rand McNally, 1975), introduce 'construct validity' as one of four aspects of research validity – the others being 'internal', 'external', and 'statistical'.

34. J. Margolis, 'Citation Indexing and Evaluation of Scientific Papers', Science, Vol. 155 (10 March 1967), 1213-19.

35. Cole and Cole, op. cit. note 30.

36. Chubin and Moitra, op. cit. note 9.

37. Chubin and Moitra, ibid.

38. Moravcsik and Murugesan, op. cit. note 1.

39. This discussion follows the approach to issues of validity of Cook and Campbell, op. cit. note 33. N.K. Namboodiri, L.F. Carter and H.M. Blalock, Jr., provide a good perspective on measurement error in Applied Multivariate Analysis and Experimental Designs (New York: McGraw-Hill, 1975).

40. Cole and Cole, op. cit. note 3.

41. The questions are raised, not to be anwered here – Cole and Cole themselves carefully address many of these – but to exemplify the threats to validity that merit consideration.