Current Comments'

How to Use Citation Analysis for Faculty Evaluations, and When Is It Relevant? Part 1

Number 44

October 31, 1983

Every year, in hundreds of universities and colleges throughout the world. thousands of academic administrators go through the painful process of evaluating tens of thousands of candidates for promotion. In most cases, their decisions are based on a faculty committee's appraisal of the candidate's teaching ability, research performance, and involvement in the scholarly community. Although this process has been used to evaluate and promote faculty for some 80 years,¹ many people now challenge the validity of this system. Some authors claim that personality and research biases play too great a role in faculty peer reviews.² Much the same is said about peer review in connection with research funding.3

Others find the faculty review process somewhat capricious. Douglas Needham, Western Kentucky University, Bowling Green, claims that the subjective nature of faculty evaluation often results in the use of inconsistent criteria to evaluate different individuals.⁴ Presumably, it's only fair that all candidates in a department be judged by consistent criteria—at least for the same position.

The economic and social significance of these evaluations should not be underestimated. Tenure, for example, involves a substantial investment for the university. By granting tenure, a university commits itself to one million dollars or more over the life of a career.² In effect, a well-researched tenure evaluation is insurance against an unwise investment. For an individual under consideration, an unfair evaluation can make an enormous difference in career goals.

This essay reviews some of the criteria used in faculty evaluations. More to the point, it explains how citation data can augment these somewhat subjective evaluations. The goals of citation analysis should be an increase in objectivity and certainly an increase in the depth of analysis. Numerous issues and techniques should be considered in citation analysis, so this essay is necessarily longer than the average *Current Comments*[®] essay. Consequently, the essay has been split in two consecutive parts.

Although I've discussed citation analysis in estimating creativity many times,^{5,6} I have generally avoided discussing its use for the kind of personal, individual evaluation implicit in promotion decisions. That is not only because citation analysis is controversial. Those of us who use citation data regularly realize that they are easily misinterpreted or inadvertently manipulated for improper purposes. Furthermore, certain techniques that don't interfere with the use of citation indexes for literature searching, or bibliometric/scientometric analysis, may introduce disproportionate problems in the use of the data for evaluating individuals.7 These concerns do not prevent us from performing large-scale studies where emphasis on particular individuals is minimized. And we are careful never to damage an individual by invidious comparisons. Instead, as reported recently,⁸ citation analysis has helped identify many deserving but unrecognized scientists.

Like it or not, and with or without any prompting from ISI®, citation analysis has become an important indicator for estimating the impact of scholarly work. Since some administrators may be using Science Citation Index[®] (SCI[®]), Social Sciences Citation Index[®] (SSCI[®]), or Arts & Humanities Citation Index " (A&HCI[™]) with less than full knowledge of what the data mean, I feel obliged to deal with their use in connection with promotion, particularly tenure, decisions. The citation investigation of a scholar, especially a prolific one, can be a complex procedure. The process involves more than a mere glance at SCI, SSCI, or A&HCI in print or online.

The ultimate evaluation involves an in-depth interpretation of each candidate's papers and books. The analysis should take into account the publishing and citing conventions of the field, the reasons why the candidate's papers are cited, and adjustments for self-citations. I make these qualifications knowing that, in many instances, one can obtain important *impressions* about individual candidates by a mere glance at a five- or ten-year cumulation of SCI or SSCI. But this first crude impression needs to be qualified by the other impressions obtained through subjective peer review.

The traditional tenure evaluation generally begins with an appraisal of the candidate by the department. The chairperson, sometimes consulting with other tenured faculty members, assesses the candidate's teaching and research abilities. In addition to supplying the committee with a curriculum vitae (CV) and bibliography, candidates may be asked to conduct self-evaluations which include evidence of teaching and research performance and community service.9 In their review of tenure evaluation systems at a number of universities, Richard P. Chait, Pennsylvania State University, University Park, and Andrew T. Ford, Allegheny College, Meadville, Pennsylvania, note that at Harvard's Graduate School of Business Administration, candidates are also permitted to specify those professors who could not dispassionately evaluate their work.¹⁰

At some universities, teaching ability is evaluated by observing candidates in the classroom and reviewing their instructional materials. In most cases, however, student evaluation forms are used. These forms permit the students to rank such teaching attributes as command of a subject, enthusiasm about the topic, and availability for student counseling. Many authors are critical of these forms, claiming they don't really mea-. sure changes in student knowledge and achievement. In their reviews of tenure and student evaluation, Peter Seldin.9 Pace University, Pleasantville, New York; Richard I. Miller,¹¹ Southwest Texas State University, San Marcos; Frank Costin and colleagues, 12 University of Illinois at Urbana-Champaign; and Ronald A. Berk,¹³ Johns Hopkins University, Baltimore, Maryland, offer guidelines on constructing and analyzing student rating instruments. Whatever methods are chosen, one cannot stress too much the importance of teaching ability and dedication. This, too, is a complex topic which extends far beyond student evaluation.

Evaluating a candidate's research performance is probably the most complicated aspect of the tenure review. Candidates are often asked to submit the names of colleagues elsewhere who can comment on the quality of their research. Letters are requested from these experts. If faculty members know others at work on research related to the candidate's, their comments may also be solicited.

Most departmental committees, armed with the candidate's bibliography, will attempt to evaluate the quantity and quality of the candidate's research. This "evaluation" sometimes consists simply of a publication count. In other cases, different weights are assigned to different types of publications. Numerous evaluative scales have been devised for this purpose. They generally state how many papers are equivalent to a book. Different weights are assigned to research papers, review articles, notes, letters, and unpublished reports.4.14-19 Some assessments call for assigning authors less than full credit for multiauthored work.^{10,16,19} Chait and Ford recommend that evaluative systems also distinguish between refereed and nonrefereed journals.¹⁰ (p. 197) However, though most prestigious journals are refereed, some are not. The weighting system at the University of San Francisco, which takes a highly quantitative approach to faculty evaluation, goes beyond most such systems. Publications are scored by their length, the journal's reputation, the number of citations received, the number of times the paper is reprinted, and whether it is given an award or discussed extensively in other papers.¹⁰

In reviewing research performance, faculty members should also read several if not all of the candidate's papers and books. This used to be an easier task. The growth of departments and individual publication output has made this much more difficult, but it is certainly necessary.

Before arriving at a decision, the departmental committee generally considers the candidate's contributions to the department, including administrative duties, advice to colleagues, and grants received. Also considered are awards, honors, and participation in scholarly societies and on editorial boards. Invitations to speak at symposia, conventions, and other universities; papers presented at meetings; and consulting work may be taken into account. Some committees may even consider interpretive articles written for the general public. Martin Meyerson, chairman of the University of Pennsylvania Foundation and a former Harvard faculty member, points out that most committees also consider whether the candidate's work will expand the department's program, give it new direction, or create bridges to other programs.²⁰

If a candidate passes muster at the departmental level, the information gathered thus far will be reviewed by a university- or college-wide committee. Membership on this committee is often confidential and limited to scholars working in fields broadly related to the candidate's specialty. At Harvard University, experts in the candidate's specialty from outside the university, and the president of the university, also serve on this committee. Meyerson says that these outside experts "are expected to ask where the subject field should be headed and not just examine past and present practices and citations. They are also expected to ask, in light of such analysis, whether or not potential candidates are apt to be in the forefront of future work."20 This committee often requests additional letters appraising the candidate. Sometimes a colleague may be asked to present the candidate's case.

After discussing this accumulated evidence, the committee will take a vote. At Harvard's Graduate School of Business Administration, they vote twice: once to evaluate the candidate separately, and once to compare the candidate with other candidates. Ordinarily, the committee informs the college dean or vice president. This administrator then makes a recommendation to the president, who, in turn, generally presents a recommendation to the university's governing board. In most cases, the administration affirms the faculty committee's decision.¹⁰ This traditional procedure for granting tenure obviously incorporates several levels of peer review. Presumably, after considering the opinions of so many colleagues, within and without the university, the administration gains a clear picture of each candidate's qualifications.

Despite the thoroughness of such reviews, the traditional evaluation process leaves much to be desired. Many authors challenge the usefulness of review letters and the politics of peer review. Others claim that counting publications ("publish or perish") or even reading the candidate's papers does not necessarily provide qualitatively useful information about his or her research qualifications. Robert K. Merton, Columbia University, New York, like so many eminent academicians, has written and reviewed many faculty evaluation letters. He points out that there's "no methodical way of assessing and comparing the estimates" provided by different evaluators unless you "know their personal scales of judgment and can read between the lines."²¹

Stephen Cole, State University of New York, Stony Brook, notes that university-wide committees may have trouble determining whether the departmental committee was being honest in assessing candidates. He writes: "There are many particularistic non-scientific factors which go into the determination of tenure decisions. There are friendships, departmental politics, pressure to fill sex or race quotas, etc. Even if decisions are made by departments using these nonscientific criteria, they generally try to mask the decisions in scientific terms. Thus, the work of a candidate will be exaggerated in its importance, or outside referees will be selected whom the Chair feels will give favorable comments."22 More than one referee for this essay felt that review committees often take negative comments more seriously than positive ones. While it is usually easy to be positive, it is also possible for one faculty member to blackball another.

Since there are thousands of journals, almost any determined faculty member can get published and thereby assemble a fairly lengthy bibliography, particularly if he or she publishes in nonrefereed or low-impact journals. Junior faculty know that the length of bibliographies will often be considered in their tenure According evaluations. to James O'Toole, University of Southern California, Los Angeles, many will publish articles that are "meretricious and banal."2 Even when members of the faculty review committee read the relevant publications, a clear impression of the candidate's qualifications may not emerge. Committee members not working in the candidate's field are unlikely to

understand the significance of the research. Even those who do work in the same specialty area may resist theories that challenge their own.²³ Or they may simply be unable to assess the importance of a candidate's work. As Cole points out, "In all sciences knowledge at the research frontier is characterized by substantial levels of disagreement and difficulty in determining which contributions will turn out to be significant."²⁴ So the faculty evaluation procedure is fraught with subjective pitfalls.

When used properly, citation analyses can help members of the faculty evaluation committee develop better informed opinions about candidates. Objective information about the usefulness of a scientist's work to the scientific community is to be welcomed. Careful analysis of the candidate's publications can confirm or contradict the conclusions of a faculty review committee. As we shall see, SCI, SSCI, or A&HCI can also facilitate the peer review process by helping administrators identify the group of scholars involved in research related to the candidate's work. These scholars will be most qualified to comment on a candidate's research. This point is often overlooked. And frequently this kind of intelligence gathering needs to be done outside the department. It isn't always possible for departmental colleagues to name the people best qualified to judge the research.

There may be good reason to do a check on one or two of the outside evaluators to find out how qualified they are to judge the candidate's work. In this case, a bibliography may not be available. SCI. SSCI. and A&HCI. and our online files such as SCISEARCH[®], Social SCISEARCH[®], ISI/BIOMED[®], and the other data bases available through the ISI Search Network can help administrators assemble such a tentative publications list quickly.25 This can provide information about the evaluator's own publication record. More importantly, by consulting citation data for the papers on this list, administrators will be able to tell if and how much this scholar's work has been used by other scholars. Not only will citation data help pinpoint the evaluator's and candidate's relevant work, it will help determine if they are recognized for similar types of research. Citation data can also help administrators determine if they are known primarily for current or past work, if they are recognized for original concepts or methods, for review articles, or for all of these.

Citation data can be useful and relevant only if one takes the time to interpret the results properly. Citations do not necessarily reflect the usefulness of research for curing disease, finding new drugs, and so on. And they tell us nothing about a candidate's teaching ability. administrative talent, and other departmental contributions. Rather, citations reflect the usefulness of research to other scientists doing related work. As such, they are what Manfred Kochen, University of Michigan, Ann Arbor, calls "lagging indicators."26 They tell us the influence a scientist has had, but not necessarily the impact he or she will have in the future. Furthermore, as explained in more detail in Part 2 of this essay, if one wants to go beyond mere impressions, one must find out not only how often, but why someone is cited. Administrators should take the time to determine normal citation rates for the field involved. The number of citations received by different types of papers, and the amount of time it takes before they reach their peak citation rates, vary from field to field. So you should compare the tenure candidate with others who are working in the same specialty. In any event, the information gathered through citation analysis should be treated as only one more, though distinctive, indicator of the candidate's influence. This indicator can be compared with others used in the traditional evaluative procedure. Citation data serve best when they complement other kinds of evidence.

The citation investigation must begin with the candidate's complete bibliography. This is, of course, generally available from candidates for tenure. However, obtaining the bibliography for an outside evaluator, or for someone you're considering for an open position, can be more difficult than many people imagine. This places the burden to assemble such bibliographies on administrators or librarians.

Most authors do maintain bibliographies, but will often include incomplete or incorrect information. For example, in a CV, authors frequently say "published jointly with." This does not identify the first author on a multiauthored paper. In such cases, administrators can check the SCI's Source Index to find out the original order in which the names appeared. This must be done before looking in the Citation Index, because papers are listed there under the first author.

In most cases, though, candidates for promotion will have taken the time to assemble complete, accurate bibliographies. This is always preferable, particularly if their names are potential homographs. When a candidate's name is a homograph, the relevant papers must be separated from those of other authors having the same surname and initials.

An important rule of thumb in the citation investigation is that one can't evaluate an individual scholar alone and out of context. You must evaluate a group. Even thoroughbred animals are evaluated in appropriate categories. Don't compare physicists with sociologists, or plant scientists with chemists. Try to find out how the candidate compares with others working in the same discipline or, preferably, the same specialty.

Several methods can be used to find the members of your candidate's "invisible college."²⁷⁻²⁹ The citation "cycling" system is often used.⁶ (p. 58) With this system, one begins by finding out which scholars are cited in the candidate's most recent papers. You then look in *SCI* to find those scholars who not only cite your candidate, but also those he has cited regularly. This can be done in one or more iterations.

To help appreciate the significance of a candidate's papers, you may wish to draw citation influence "maps" indicating relationships between papers and authors in the field. Such a map is illustrated in Figure 1. It was compiled for an essay in which we used citation analysis to examine the work of Lester R. Aronson, a scientist involved in an antivivisection controversy.²⁷

A single citation link between two authors is usually a weak indicator that they are working in the same field. The uncertainty can be eliminated by including in your list or inventory only those authors who have referred to the candidate's work in at least two publications. Furthermore, as in Figure 1, we only included those authors who were cited by the candidate at least three times. In this manner, 21 people were identified as members of this author's invisible college. The citation connections between the 22 authors are also shown in Figure 1. There is a growing literature on techniques for mapping fields. Peter Lenk, University of Michigan, offers a method for using peer nominations to create these maps.²⁸

Depending upon the depth of the analysis required, these procedures may or may not be time-consuming. But I never said a citation investigation would be easy. Fortunately, a more direct source for identifying the most active invisible colleges is available. The technique of co-citation analysis has been used to classify literature in data bases like ISI/BIOMED, ISI/CompuMath[®], and ISI/GeoSciTech[™]. Using the entire multidisciplinary SCI data base for one year, we have also identified over 2,000 subspecialties, or research fronts, for

Figure 1: Historiograph based on research on sensory and hormonal influences on cat sexual behavior. Each node represents a paper or group of papers by the same authors; larger nodes represent more than one paper. Lines between nodes represent citations; a single line may represent multiple citations. The most recent contributions appear toward the bottom.



Index to Scientific Reviews $(ISR^{**}).$ These specialties are identified when a group of current papers cite one or more core papers for that topic.³⁰ The authors of core documents are generally active scholars in the area involved. If your candidate's papers are among the core, you can more easily study in detail the microstructure of that field. The number of current papers published will give you an idea of the number of people working in this research area. If the time lag between the core and current papers is less than three or four years, then you know that you are dealing with a rapidly changing field.

Howard D. White and Belver C. Griffith, Drexel University, Philadelphia, explain how you can use ISI's online data bases to create your own co-citation clusters.²⁹ Their method involves identifying recognized *scholars* in a field, and finding out who has co-cited them.

Cole suggests still another technique for selecting an appropriate group with whom to compare the candidate. He proposes that candidates be compared with faculty members who have been promoted or granted tenure at equal caliber departments in the last several years. Sharp discrepancies between the number of citations received by the candidate and these other faculty members should be questioned.²²

The invisible college members identified through citation networking or mapping can be asked for their judgments of the quality of the candidate's work. Since they are working in the same or a similar specialty, they will be best situated to assess the significance of the work.

The invisible college will also provide evaluators with a pool of colleagues with whom a candidate can be compared. A number of factors should be considered in this comparison. The first, and perhaps most important, point is that small differences in the number of citations are generally not statistically significant. It's absurd to conclude that a candidate with 15 citations is more influential than a candidate with ten. This problem may be particularly acute in tenure evaluations, since candidates may not have been publishing long enough to have acquired a substantial number of citations. For this reason, Cole believes that citation data are more useful in making decisions about promotions from associate to full professor. He has found that "at this stage in a scientist's career there should be some significant trace of the utility of their work in citation patterns."²² In other words, except in rare cases of precocious talent, it may be inappropriate to use citation analysis when evaluating candidates for lower ranking positions.

Assistant professors are usually hired for two three-year terms. American Association of University Professors regulations require a promotion decision within seven years. Most universities make the decision in the candidate's sixth year. The dossier that goes out to external reviewers is compiled in the summer between the assistant professor's fifth and sixth years. Many dossiers contain 25 percent or more preprints or unpublished papers. Many who reviewed this essay indicated that the best original papers of candidates are published within a year or two of the evaluation, so they reach the expert in preprint form. There is generally not enough time for those papers to have been cited. So at a time when citation data would be most useful, it is not yet available. The tough decision must often be made on the basis of reading the work and judging its merit without benefit of formal indications that the scholarly community recognizes the value of the work.

Roald Hoffmann, Cornell University, Ithaca, New York, was one of many people who reviewed this essay. He suggested an important experiment. Select a field and a year when a dozen assistant professors began their careers. For example, select chemists starting in 1970. Select six who were promoted to tenure in 1976 and six who were not. Perform a citation analysis based on 1970-1975 data. Then repeat the analysis for 1978-1983. Is there a correlation between citations prior to tenure decisions, the tenure decision, and the longer term analysis?

While limited to a smaller group of candidates one could also use our annual compilations of most-cited papers to identify a *young* academic's work. So in each year we would be identifying candidates for tenure provided we were able to eliminate collaborations that depended heavily on senior investigators. And it is important to realize that age alone cannot be used reliably because people become associate professors over a tenyear span, especially in the humanities.

Another point to remember is that more established scientists may have accumulated more citations than less experienced scientists. This is partly because an older scholar's papers may have been in circulation longer. For this reason, it is preferable that scientists be compared with relevant invisible college members with roughly the same number of years of experience.³¹ This is preferable to comparing scholars of the same age, since the age at which scientists begin their careers varies widely.

Evaluators should also note that scientists who've published with a team may be more prolific. Presumably, they may accumulate more citations.³² One should also carefully observe those who regularly publish the "least publishable unit." Such "fragmentation" occurs where authors publish numerous short papers, rather than one consolidated work.³³ On the other hand, duplication, or publishing several papers on the same, or overlapping, data, is possible.

Finally, although some self-citation is reasonable and expected in science, some evaluators are especially sensitive to excessive self-citation. Edward Anders, Enrico Fermi Institute, University of Chicago, checks the candidate's bibliography against a list of papers that cite the candidate.³⁴ In this manner, he can find self-citations from papers on which the candidate was a secondary author. This is necessary because, as mentioned earlier, papers are listed by first author in SCI. In some cases, it may be more relevant to examine the number of citing authors, or papers, rather than citations per se. This is particularly important in the social sciences and humanities, where a single author may be cited repeatedly in one paper. In our analysis of the most-cited authors in the arts and humanities, this became particularly relevant.35

(To be continued.)

My thanks to Joan Lipinsky Cochran and Amy Stone for their help in the preparation of this essay.

REFERENCES

- Metzger W P. The history of tenure. (Metzger W P, O'Toole J & Glazer P M.) Tenure. Washington, DC: American Association for Higher Education. 1979. p. 3-12.
- 2. O'Toole J. The poverty of historicism: a critique of Walter Metzger's defense of tenure. (Metzger W P, O'Toole J & Glazer P M.) Tenure.
 - Washington, DC: American Association for Higher Education, 1979. p. 13-4.
- 3. Cole S, Cole J R & Simon G A. Chance and consensus in peer review. Science 214:881-6, 1981.
- 4. Needham D. Improving faculty evaluation and reward systems. J. Econ. Educ. 13(1):6-18, 1982.
- Garfield E. Citation measures used as an objective estimate of creativity. Essays of an information scientist. Philadelphia: ISI Press, 1977, Vol. 1, p. 120-1.
 - (Reprinted from: Current Contents (34):4-5, 26 August 1970.)
- 6. -----Citation indexing—its theory and application in science, technology, and humanities. Philadelphia: ISI Press, 1983. 274 p.
- 7. Idiosyncrasies and errors, or the terrible things journals do to us. *Current Contents* (2):5-11, 10 January 1983.
- ----- How sweet it is---the ACS Patterson-Crane Award. Reflections on the reward system of science. Current Contents (30):5-12, 25 July 1983.
- 9. Seldin P. Successful faculty evaluation programs. Crugers, NY: Coventry Press, 1980. 182 p.

- 10. Chait R P & Ford A T. Beyond traditional tenure. San Francisco, CA: Jossey-Bass, 1982. 291 p.
- 11. Müler R I. Evaluating faculty performance. San Francisco, CA: Jossey-Bass, 1972. 145 p.
- 12. Costin F, Greenough W T & Menges R J. Student ratings of college teaching: reliability.
- validity, and usefulness. Rev. Educ. Res. 41:511-35, 1971.
- Berk R A. The construction of rating instruments for faculty evaluation. J. High. Educ. 50:650-69, 1979.
- Crane D. Scientists at major and minor universities: a study of productivity and recognition. Amer. Sociol. Rev. 30:699-714, 1965.
- 15. Endler N S. Beyond citation counts: developing research profiles.
 - Can. Psychol. Rev.—Psychol. Can. 19:152-7, 1978.
- 16. Manis J G. Some academic influences upon publication productivity. Soc. Forces 29:267-72, 1951.
- 17. Meltzer B N. The productivity of social scientists. Amer. J. Sociol. 55(1):25-9, 1949.
- Lawani S M. Citation analysis and the quality of scientific productivity. Bioscience 27:26-31, 1977.
- 19. Price D J D. Letter to editor. (Multiple authorship.) Science 212:986, 1981.
- 20. Meyerson M. Personal communication. 19 August 1983.
- 21. Merton R K. Telephone communication. 17 May 1983.
- 22. Cole S. Personal communication. 9 August 1983.
- Osmond D H. Malice's wonderland: research funding and peer review. J. Neurobiol. 14(2):95-112, 1983.
- 24. Cole S. The hierarchy of the sciences? Amer. J. Sociol. 89(1):111-39, 1983.
- 25. Garfield E. How to use Science Citation Index (SCI). Current Contents (9):5-14, 28 February 1983.
- 26. Kochen M. Telephone communication. 21 September 1983.
- Garfield E. Citation analysis and the anti-vivisection controversy. Part II. An assessment of Lester R. Aronson's citation record. *Essays of an information scientist*. Philadelphia: ISI Press, 1980, Vol. 3. p. 316-25. (Reprinted from: *Current Contents* (48):5-14, 28 November 1977.)
- Lenk P. Mappings of fields based on nominations. J. Amer. Soc. Inform. Sci. 34(2):115-22, 1983.
 White H D & Grifflith B C. Authors as markers of intellectual space: co-citation in studies of science,
- technology and society. J. Doc. 38:255-72, 1982.
- 30. Garfield E. ABCs of cluster mapping. Parts 1 & 2. Most active fields in the life and physical sciences in 1978. Essays of an information scientist. Philadelphia: ISI Press, 1981. Vol. 4, p. 634-49, (Reprinted from: Current Contents (40):5-12, 6 October 1980 and (41):5-12, 13 October 1980.)
- 31. Price D J D. Telephone communication. 12 August 1983.
- 32. Price D J D & Beaver D B. Collaboration in an invisible college. Amer. Psychol. 21:1011-8, 1966.
- 33. Broad W J. The publishing game: getting more for less. Science 211:1137-9, 1981.
- 34. Anders E. Personal communication. 15 July 1983.
- Garfield E. Most-cited authors in the arts and humanities, 1977-1978. Essays of an information scientist. Philadelphia: ISI Press, 1981. Vol. 4, p. 238-43. (Reprinted from: Current Contents (32):5-10, 6 August 1979.)