

Citation Ranking Versus Peer Evaluation of Senior Faculty Research Performance: A Case Study of Kurdish Scholarship

Lokman I. Meho,* and Diane H. Sonnenwald

School of Information and Library Science, University of North Carolina at Chapel Hill, CB#3360, 100 Manning Hall, Chapel Hill, NC 27599-3360. E-mail: {mehol, dhs}@ils.unc.edu

The purpose of this study is to analyze the relationship between citation ranking and peer evaluation in assessing senior faculty research performance. Other studies typically derive their peer evaluation data directly from referees, often in the form of ranking. This study uses two additional sources of peer evaluation data: citation content analysis and book review content analysis. Two main questions are investigated: (a) To what degree does citation ranking correlate with data from citation content analysis, book reviews, and peer ranking? (b) Is citation ranking a valid evaluative indicator of research performance of senior faculty members? Citation data, book reviews, and peer ranking were compiled and examined for faculty members specializing in Kurdish studies. Analysis shows that normalized citation ranking and citation content analysis data yield identical ranking results. Analysis also shows that normalized citation ranking and citation content analysis, book reviews, and peer ranking perform similarly (i.e., are highly correlated) for high-ranked and low-ranked senior scholars. Additional evaluation methods and measures that take into account the context and content of research appear to be needed to effectively evaluate senior scholars whose performance ranks relatively in the middle. Citation content analysis data did appear to give some specific and important insights into the quality of research of these middle performers, however, further analysis and research is needed to validate this finding. This study shows that citation ranking can provide a valid indicator for comparative evaluation of senior faculty research performance.

1. Introduction

There are two fundamental purposes for evaluating faculty research performance: first, to identify areas for improvement in the quality of their research and, second, to

help committees in funding, hiring, and making promotion and tenure decisions. Basically, all research performance evaluation systems revolve around the concept of quality. Although quality in this sense is generally defined as a measure of the extent to which an idea or an author has contributed to the progress of knowledge, “ultimately, it is always the scientific community . . . who will have to decide in an ‘inter-subjective’ way about quality” (van Raan, 1996, p. 398). Until recently, the general practice for evaluating the research performance of faculty members was to rely on two interrelated criteria: the opinions of colleagues—in the form of peer review, evaluation, judgment, or ranking—and the individual’s list of publications. Evaluations based on the former criterion are almost universal in terms of the method used for data collection, e.g., statements of support. Evaluations based on the latter criterion range from simple evaluation systems such as publication counts, to systems that devise certain evaluative or weighting scales for different kinds of publications—books, chapters in books, articles in refereed and nonrefereed journals, conference papers, etc. In many cases, publication length and the reputation, or rank, of the place of publication are taken into consideration as well.

Certainly, a long list of publications that includes books or articles in refereed journals is a putative indicator of research productivity, but it neither conclusively says anything about the quality of the work being published nor about its contribution to, or influence on, a body of knowledge. Similarly, evaluations based on “peer review” are widely perceived as an acceptable and respected academic norm for evaluating the research performance of faculty members. However, these evaluations suffer significantly from the level of knowledge, integrity, and research biases of the evaluators and committee members (Chubin & Hackett, 1990; Garfield, 1983; King, 1987). In connection with this, concern has been expressed over the need for additional complementary indicators of research performance. After all, the use of several sources of information or

* To whom all correspondence should be addressed.

Received February 24, 1999; revised June 3, 1999; accepted June 3, 1999.

© 2000 John Wiley & Sons, Inc.

multiple indicators not only help in correcting misleading assumptions, it also throws into relief the theoretical notions which are used or implied in interpreting empirical material (Mulkey, 1974). Along the same lines, Baird and Oppenheim (1994), in their review of citation-based studies, conclude that:

[T]here is not, and never can be, one single measure of the value of information that will be universally acceptable. However, there are a number of measures that might, in combination, lead to some sort of index of the value of a piece of information, an individual's contribution, or a collection of information (p. 13).

The advantages of using multiple indicators (or what is termed as "the methodology of converging partial indicators") are discussed in detail by Irvine and Martin (1983).

Among the various methods that have been proposed and used to complement the traditional research performance evaluation measures, citation analysis was and still is one of the most widely used (Chubin & Hackett, 1990; Martin, 1996; Thomas & Watkins, 1998). Citation analysis—exemplified in this study by citation ranking or citation counts—claims to provide researchers with an effective indicator for assessing not only the research performance of individual authors, but also for assessing the relative quality of papers, journals, programs, etc. (Garfield, 1983). This paper investigates this claim and examines the relationship between citation ranking and peer evaluations in the context of Kurdish studies. More specifically, this paper assesses the quality of research produced by a group of senior Kurdologists¹ on the basis of three peer evaluation data sets—citation content analysis, book reviews, and peer ranking—and then analyzes the relationships between the three data sets and citation ranking. Under the assumption that peer evaluation is a valid and the most accepted indicator of scholarly research performance, the main questions examined here are:

- To what degree does citation ranking correlate with data from citation content analysis, book reviews, and peer ranking?
- Is citation ranking a valid evaluative indicator of research performance of senior faculty members?

These questions are examined in the context of Kurdish studies. In particular, the research performance of five senior social science Kurdologists is examined. All of these Kurdologists are currently active faculty members in Western academic institutions. The field, Kurdish studies, was selected because of its recent publishing trend as well as because of the increasing number of academic scholars who

specialize in the field. Research on the Kurds has been conducted by national, regional, as well as international scholars for over a century. However, the 1990s in particular, witnessed a boom in Kurdish literature. For example, according to Online Computer Library Center's (OCLC) WorldCat, of the 2,800 20th century items about the Kurds recorded in its database through April 1999, 1382 (or approximately 50%) of them were published between 1990 and 1999. In other words, the number of works being published on the Kurds in the last nine years is almost equal to the total number of works published about them in the previous 90 years. This recently produced literature is so large it raises interest in analyzing its quality and the contribution to, or influence on, the body of knowledge its authors are making.

To examine the senior Kurdologists' research performance and examine the relationship between citation ranking and peer evaluation data, the five study participants were ranked on the basis of the number of citations they obtained through 1998; quality of the citations as determined by citation content analysis; quality of books they published as determined by a content analysis of book reviews; and peer ranking. Citation ranking is then compared to the peer evaluation data sets (i.e., citation content analysis data, book review data, and peer ranking). Results show that normalized citation ranking (total number of citations over the number of years since publishing the first work on the Kurds after dissertation) is a valid measure for evaluating research performance of senior faculty, and correlates highly with citation content analysis data, book review data, and peer ranking for high-ranked and low-ranked senior scholars.

2. Limitations of the Study

It could be argued that a major limitation in this study is the small size of its population, something which might raise some questions on the generalizability of the findings. However, it should be emphasized that after a painstaking search and many efforts to include as many subjects as possible, only five could be identified. A population of five is small, but it reflects most real life situations when it comes to funding, hiring, and making promotion and tenure decisions. Given that the number of Kurdologists is on the rise, the number of participants might be expanded at a later date so that future findings could more reflect the appropriateness and validity of citation ranking for research evaluation. Alternatively, a different and larger group of scholars could be selected for that purpose. Another possible limitation in this study could be the method employed in normalizing citation counts and thus citation ranking. Citation counts were normalized by dividing the total number of citations over the number of years since publishing the first work on the Kurds after dissertation. Certainly, there are many faculty today who published on the Kurds long time before they finish their doctoral dissertations. However, it can also be argued that it may take years before many faculty publish after completing their dissertations. Many

¹ A Kurdologist is defined here as any researcher: (1) whose list of scholarly publications on the Kurds outnumbers that of other topics; (2) who writes on the Kurds only or; (3) who publishes significantly on them (for example, has published a book or is publishing one scholarly work or more on the Kurds every year).

others may shift research interests in the middle of their careers as is the case with two of the study participants. In any case, of the five study participants, only one did publish before completing his dissertation, but that was only one year before and hence had no impact on the results.

Despite the aforementioned limitations, this study still has a significant value, represented in its use of multiple indicators for assessing research performance of senior scholars, in the insights it provides and questions it raises for both research evaluation committees and bibliometric scholars, and in the example it provides of how multiple evaluative data sets can be used in real life situations when hiring, promotion, and tenure committees are faced with the task of evaluating senior candidates. The study also presents several questions that warrant further research.

3. Literature Review

3.1 Citation Ranking, Peer Evaluation, and Research Assessment

The use of citation ranking, or citation counts,² for evaluating research quality is based on the assumption that citations are a form of giving credit to or recognizing the value, quality, significance, or impact of the authors' work (Cole & Cole, 1967, 1968). For a detailed discussion on this point, see Gilbert (1978), Smith (1981), van Raan (1996), and White (1990). Many have argued for and against the use of citations for research quality assessments. While the proponents argue that this method is an indispensable support tool for traditional evaluative measures (Cronin & Overfelt, 1994; Garfield, 1979, 1983; Glanzel, 1996; Koenig, 1982, 1983; Kostoff, 1996; Lawani & Bayer, 1983; Narin, 1976; Narin & Hamilton, 1996; van Raan, 1996, 1997), critics claim that it has some serious problems or limitations that impact its validity (MacRoberts & MacRoberts, 1986, 1989, 1996; Seglen, 1992, 1998).

Important limitations reported in the literature focus on the nature of citation counts and citation databases. Citation counts give no clue why a work is being cited, are field-dependent, and may be influenced by time, number of publications, access to, or knowledge of the existence of, needed information, visibility and/or professional rank of the authors, self-citations, or citations from colleagues. Citation databases provide credit only to the first author, cover mainly English journal articles published in the United States, are not comprehensive in coverage, and have many technical problems such as synonyms, homonyms, clerical errors, and limited coverage of the literature.

The limitations of citation count studies are typically caused by the misuse of the method and misinterpretation of the results rather than the problems that are inherent in the citations and citation databases. Garfield (1979, 1983), in

his seminal works on citations, warned against the misuse of citations and, at the same time, explained when to use them and for what purposes. Smith (1981) did the same too. Perhaps more importantly, she indicated that unless the limitations, discussed above, are taken into consideration in the design of a study and in the interpretation of the results, invalid conclusions would be made. In summarizing the validity of citation counts for evaluation purposes, Koenig (1983) states that:

Despite the ambiguities of citation practices, the difficulties of ascertaining why a paper is or is not cited, and the potential malpractices in citing, considerable evidence has been accumulated to suggest that citations do indeed provide an objective measure of what is variously termed 'productivity,' 'significance,' 'quality,' 'utility,' 'influence,' 'effectiveness,' or 'impact' of scientists and their scholarly products. (p. 61)

Another method used to evaluate scholars is peer ranking. Three primary criticisms of peer evaluation methods have been discussed in the literature. First, respondents may lack sufficient knowledge on which to base judgments on a particular peer. Second, judgments are often determined by unstated criteria that are probably not consistent from person to person. Third, different peers in different cognitive and social locations may evaluate a given scientific contribution (and therefore its author) rather differently (Martin, 1996). In short, evaluations based on peer ranking are influenced significantly by the level of knowledge and research biases of the evaluators. For a detailed and authoritative review of the overall peer review system in the United States (and beyond), see Chubin and Hackett (1990).

Studies that report both the validity of citation counts in research assessments and the positive correlation between them and peer evaluations have been discussed and reviewed by many, including Baird & Oppenheim (1994), Kostoff (1996), Narin (1976), Narin & Hamilton (1996), and Smith (1981). The following studies are examples of papers discussed in these reviews.

In his pioneering work, Clark (1957) analyzed the relationship between different measures of eminence for 566 highly visible individual psychologists. The variables used to measure the eminence of the subjects were research productivity, peer votes, American Psychological Association office held, citations received from journals, and citations received from the Annual Review of Psychology. The intercorrelations of the measures of eminence for the highly visible psychologists showed that the highest correlation is between the number of votes each psychologist received as a "significant contributor" to the field and citations received from journals ($r = 0.67$). Myers (1970) examined lists of the most cited authors in psychology with 15 independent measures of eminence (e.g., being listed in American Men of Science, receiving scientific contribution awards, election to presidency of the American Psychological Association, etc.) and found that citation count is a good index of a

² Citation ranking and citation counts are used interchangeably, depending on the context.

scientist's esteem. Virgo (1977) found citation count to be a consistent and accurate predictor of important scientific papers as determined by her judges. The meaning of "importance" was left to her judges to determine individually from a series of questions asked about each article under examination. Narin (1976) provides one of the most comprehensive works validating citation analysis. He reviews 24 studies showing that citation counts, as well as other bibliometric measures, correlate in the range of $r = 0.5$ to 0.8 with various other rankings of eminence. Lawani and Bayer (1983) found in a study of 870 cancer research papers that even after controlling for factors like self-citations, language, and country of authorship, highly rated papers were more highly cited than average papers.

In summary, studies show that despite the criticisms of citation count studies, research examining the correlation between citation counts and peer judgment have indicated a positive relationship between the two evaluation methods. This study builds on this research by investigating the relationships among citation ranking, citation content, book reviews, and peer ranking.

3.2 *Book Reviews and Research Evaluation*

There are mixed opinions about book reviews within the academic community. Some perceive them as works of low scholarly status (Riley & Spreitzer, 1970), others find them as significant assessments of the quality of published work and thus may be useful in making decisions about hiring, promotions, and salary increases of scholars (Ingram & Mills, 1989; Snizek & Furham, 1979). For example, if a candidate for promotion has published a book, the members of the promotion committee may read at least some of the available reviews and examine whether these reviews provide some knowledge about the quality of the candidate's published work. If so, the book reviews may be used to help evaluate the candidate.

The major problem associated with book reviews (as reported in the literature) is their evaluative content. For example, Champion & Morris (1973) and Moxley (1992) indicate that many reviewers tend not to give negative comments on books because the authors they review will know their identities. Without providing evidence of this, they claim that an implication of this is that book reviewers avoid or fail to provide sound criticisms of the works reviewed for fear of adverse reaction from their colleagues and for hope of receiving similar treatment from them, should the opportunity arise. Other problems with using book reviews as an evaluation measure include: the lack of consensus on standards, personal biases of the reviewers for, or against, the authors of the work they review, knowledge level of the reviewers, and editorial biases (Bornstein, 1991; Glenn, 1978; Riley & Spreitzer, 1970; Schwartz, 1989; Stieg Dalton, 1995).

Different studies have reached different conclusions concerning the evaluative content of book reviews. Champion and Morris (1973) found that in sociology journals less than

20% of all book reviews were negative. Casey (1985) found similar results in his analysis of multiple reviews for one hundred titles in American history. Natowitz and Carlo (1997) reported that books that were assessed as outstanding or favorable by one journal were assessed so by two other journals in an overwhelming majority (71% to 86% agreement). On the other hand, Bilhartz (1984) found that only slightly more than half of the books in his sample of history journals were favorably reviewed. Snizek and Fuhrman (1979) discovered that there was only 40% agreement in the content of reviews of the same books among three top sociology journals.

Hirsch, Kulley, and Efron (1974) examined the relationship between the evaluative content of book reviews and the professional status of the reviewers in political science, philosophy and economics. They found that high-status authors were more critically reviewed than low-status authors. They also found that the higher the status of the reviewer, the more favorable the review (based on Lindholm-Romantschuk, 1998). In an earlier study of the influence of professional status upon book reviewing in sociology, Taylor (1967) found that low-status reviewers write more favorable reviews, especially of books written by high-status authors. On the other hand, high-status reviewers were found to write favorable reviews of low-status authors, and unfavorable reviews of high-status authors.

While some argue that book reviews do not show enough independence and consensus for the reviews to be a good basis for research evaluation and personnel decisions, it has been suggested by others that they may still be used for these purposes if properly employed. This is particularly true since other bases of peer evaluation methods are often no more satisfactory (Glenn, 1978). In summary, the literature suggests that book reviews may be used, but with "extreme caution." As Glenn suggests, "a totally inappropriate procedure is to derive a general rating from a review and to use that rating mechanically in the evaluation process. . . . Rather, only specific criticisms and points of praise should be considered" (p. 255). Accordingly, it is presumed that this task can be achieved through clear and accurately defined categories of positive and negative comments in the reviews, as undertaken in the present study.

3.3 *Citation Content Analysis and Book Review Content Analysis and Research Evaluation*

To assess the quality of citations and describe their roles in a given body of literature, citations should be content analyzed. Similarly, the content of book reviews should be analyzed in order to identify and classify positive and negative comments in them. The burden of this task is lamentable, but the payoff in knowledge may justify the expenditure of resources.

Content analysis consists of extracting and evaluating the occurrences of the manifest and latent content of a body of textual material (e.g., a book or an article) in a systematic and generally quantitative manner. This extraction and eval-

uation is done in order to uncover key symbols and themes and to compare them to one another through the use of different types of analysis (Allen, 1989; Busha & Harter, 1980). Allen and Reser (1990) distinguish between two kinds of content analysis: classification and elemental. The former method, which is used in the present paper, "assigns documents (or other means of communication) to classes or categories to quantify one or more of their characteristics" (p. 253). This method is described in detail in Berelson (1952) and Holsti (1969). The elemental method is based on the recording of word or word group frequencies from these documents, and quantifying them as term frequency or term weighting data. This method is described in Hicks, Rush, and Strong (1985).

Content analysis, as any other research method, has its strengths and weaknesses. Busha and Harter (1980) have advised that in order for content analysis to produce good results as a system of measurement, developing an appropriate content classification scheme is an important step. They and Allen and Reser (1990) have also advised that the categories chosen should be exhaustive, mutually exclusive, clearly and accurately defined, and conceptually valid in relation to the research question. The classification and measurement of data must be undertaken with objectivity, exactness, and rigor as well. It should be noted that the assessment of the quality, context, and underlying motives of citations and book review content involves a large degree of personal judgment as well as an in-depth knowledge of the subject matter. Also, the manifest roles of citations (and comments in book reviews) differ from one field to another and therefore universal classification systems seem inappropriate (Liu, 1993; Spiegel-Rosing, 1977). For this reason, two special classification systems were developed for this study: one for the citations and another for the book reviews. While the system developed for the citations is based on both the literature examined as well as categories defined and adopted from earlier studies, the classification system developed for the book reviews is unique. These classification systems are defined below.

4. Research Methods

4.1 Study Population

The subjects included in this study were limited to Kurdologists who are currently senior faculty members (full and associate professors) teaching at Western academic institutions and who have been publishing primarily in English on the political and historical discourses of the Kurdish question. All these criteria were used to generate a population that is as homogeneous as possible. The homogeneity of the population is essential not only to make valid conclusions and interpret results correctly, but also to test how citation ranking performs in such cases.

An additional selection criterion focuses on book reviews. Because the present study relies on book review data, the senior Kurdologists must also have published at least

one book on the Kurds before 1993, and their books must be comprehensively reviewed (more than 250 words for each review) in at least three different places. The year 1993 was selected for two reasons. First, it may take up to 5 years before all reviews for a book may be published. Second, Kurdologists tend to use books more frequently than any other types of published works (Meho & Haas, in preparation). Accordingly, authors who published books before 1993 will probably be more visible to other scholars than authors who published books, for example, in 1997.

Study participants were identified and located using several sources, including specialized directories³ and research centers⁴ (for identifying and locating faculty Kurdologists), OCLC's WorldCat (for identifying authors of books), and biographical information found in published books and articles (additional sources for identifying and locating faculty Kurdologists). These sources identified 27 individuals, five of whom satisfied our selection criteria. To ensure that all possible subjects were included, the five study participants identified in the initial phase were asked to identify other colleagues who would possibly meet the criteria. Of the 11 new names suggested, none met all the criteria.

4.2 Data Collection

In this study, four groups of data were compiled: study participants' lists of publications, citations, book reviews, and peer ranking.

4.2.1 Publication data

All study participants were asked to provide a list of their publications on the Kurds to guarantee a comprehensive data set. Although the majority of the participants did provide the requested information (four out of five), searches were made on multiple databases⁵ and the results were matched with the lists provided. This search was necessary for two main reasons: to check whether the participants reported all published literature accurately and to compile a list of publications for the participant who did not provide such information. Approximately 95% of the participants' publications were collected and analyzed.

4.2.2 Citation data

Citation data on the 5 participants were collected from the scholarly journal articles, chapters, and conference papers they published. To diversify the sources of the citations and to compile as many citations obtained by the partici-

³ For example, the Middle East Studies Association of North America: Directory of Members (1995).

⁴ For example, Badlisy Center for Kurdish Studies (Tallahassee, FL).

⁵ These databases include: Historical Abstracts, International Political Science Abstracts, PAIS, Political Science Abstracts, Social Sciences Citation Index, and Sociological Abstracts.

TABLE 1. Number of citations for each participant.

Participant	No. of citations in SSCI (1998) and AHCI (1998)	Total no. of citations*	Total no. of citations excluding self-citations*	% of self-citations
A	35	248	183	26.0
D	22	110	71	35.0
C	19	91	59	35.0
E	8	44	33	25.0
B	8	42	32	24.0

* Source: 350 scholarly works examined manually. Includes citations from SSCI and AHCI.

pants as possible, data was also collected from the non-book materials published by 33 other Kurdologists⁶ as well as from the Social Sciences Citation Index (SSCI, 1998) and Arts and Humanities Citation Index (AHCI, 1998). The SSCI and AHCI were not used exclusively because of their limited coverage of related literature. The publication information of the 33 candidates was compiled using the same method used for the participants. As for the SSCI and AHCI, citations for the names of the participants with all their possible variations were searched and the duplicates were removed. Differences in results between electronic sources and manual examination are illustrated in Table 1. All-in-all, 350 papers from 128 different authors were collected and manually examined to identify citations. Although the books published by the participants were available, they were not used as sources of citations because they were found to be primarily compilations of previously published articles.

Two sets of data were derived from the citations compiled: citation counts and reasons for citing. Organizing the data for the first set was a simple straightforward task; the number of citations was counted for each participant. Organizing the data for the second set, however, required that the content of each citation be analyzed. Overall, from the 350 scholarly works examined, 378 citations were found to be related to the participants (excluding self-citations).⁷ These were analyzed and coded according to six criteria⁸: paying homage (recognition, acclaim, or praise) to peers and/or commending their work, recommending other readings, supporting a personal idea/argument, value-free reference attitude, criticizing a work, and disputing or disclaiming a work or an idea of another author. Examples of these criteria are illustrated in Table 2.

For the purposes of this study, the first four criteria were considered as positively-oriented or neutral citations and the

last two as negatively-oriented citations. As will be shown in the discussion section, this classification scheme made correlation analysis between citation counts and citation content data possible. A similar scheme was followed for evaluation data derived from book reviews.

One of the authors and three other coders applied the criteria to a sample of citations ($n = 35$). The results were compared in order to standardize and refine the wording and use of the categories. The level of agreement between the four coders before refining the wording of the categories was found to be 89% (96% between three coders). Once definitions and guidelines were agreed upon, the author coded and verified the entire data set.

4.2.3 Book review data

Scholarly reviews of the study participants' books were compiled using information provided by the participants as well as from using different databases, including: Arts & Humanities Citation Index, Book Review Digest, Book Review Index, Index Islamicus, and Social Sciences Citation Index. All-in-all, 44 reviews were identified and analyzed. The range of reviews found for each book was 3 to 12.

TABLE 2. Categories defined for citations.

Categories of citations	Examples quoted from the citations
Paying homage (recognition, acclaim, praise) to peers and/or commending their work	For a/an {valuable, great, impressive, fascinating, well-documented, excellent, etc.} work, see . . .
Recommending other readings	For further details on this topic, see . . .
Supporting a personal idea/argument	This is similar to what [author x] has argued. An argument also presented/discussed by . . .
Value-free reference attitude	[A citation without any discussion]
Criticizing a work	X ignores the fact that . . .
Disputing or disclaiming a work or an idea of another author	This claim is definitely not true . . .

⁶ All-in-all, 38 Kurdologists were identified in this study, but only 5 satisfied our criteria.

⁷ Self-citations were excluded from the analysis primarily because we are using the contents of citations to generate one set of peer research evaluation data. Usually, an author either praises his/her earlier works or, at least, doesn't criticize them.

⁸ These criteria were derived from examining the study participants' publications, as well as from criteria identified in Baird and Oppenheim (1994), Garfield (1979), Liu (1993) and Peritz (1983).

TABLE 3. Categories defined for book reviews.

Categories	Examples of positive comments quoted from the reviews	Examples of negative comments quoted from the reviews
Originality and innovativeness	Excellent contribution, provides useful/interesting information	Routine, not original, repetitive, dull
Quality of analysis	Strong, rigorous, disciplined, rational, objective, systematic, complete	Poor, erroneous, weak, incomplete, based on factual errors, distorted
Objectivity	Unbiased, objective	Biased, subjective
Research design/method	Strong	Poor
Quality of sources used	Primary, accurate, objective	Over-use of secondary sources, use of biased/inaccurate sources
Readability	Easy to read/follow, well-structured	Confusing, irritating, too many typos, repetitive
General attributes	Authoritative, excellent, fascinating, impressive, scholarly, succinct, timely, well-balanced, useful, well-researched, valuable, recommended	Nonscholarly work/research

The reviews were manually examined and evaluative comments (e.g., quality of sources used, objectivity or neutrality, and quality of analysis) were classified according to seven broad categories: originality and innovativeness, quality of analysis, objectivity, research design/method, quality of sources used, readability, and general attributes. Within each category, the comments were then grouped as positive or negative. Examples of positive and negative comments are illustrated in Table 3. Examples provided in the reviews to justify (or explain) the comments made were not counted.

One of the authors and three other coders applied the criteria to a sample of four book reviews, then the results were compared in order to standardize and refine the wording and use of the categories. Interestingly, all coders inferred exactly 17 evaluative cases from each of the book reviews. The level of agreement between the four coders before refining the wording of the categories was found to be 82% (and 96% between three coders). Once definitions and guidelines were agreed upon, the author coded the entire data set. Data obtained from content analyzing the book reviews were normalized according to the number of reviews each book obtained.

4.2.4 Peer ranking data

Twenty-eight Kurdologists—26 faculty and two nonfaculty—including the study participants, were sent a brief questionnaire asking them to rank the top 20 Kurdologists that they believe have most contributed to, or influenced, the area of Kurdish research. The questionnaire provided a list of 50 names including the participants' names, and respondents were asked to add people to their list of the top 20 if they did not find their selection among the 50. Respondents were also asked not to rank themselves. The 50 names were selected based on their publications record on the Kurds (data derived from Meho, 1997 and International Society Kurdistan, 1968) and on the number of citations they received (according to SSCI and AHCI). Of the 50 names

suggested, 47 were given at least a rank by one or more of the 17 (or 60%) people who replied to the questionnaire. Sixteen non-listed individuals were also ranked among the top 20.⁹ The total number of all ranked people was 63, including the study participants.

Because the respondents were asked to rank only 20 individuals, a statistical procedure was followed to generate a unique ranking score for each ranked person. Specifically, for each response, every top ranked person was given 20 points, the person ranked second was given 19 points, the person ranked third was given 18 points and so on until the person ranked 20th. The last ranked person got 1 point. It should be noted that one respondent ranked only 18 persons and another respondent ranked 20 but divided them into two groups: academic (14) versus nonacademic (6). Points for the persons named in these cases were calculated starting from the top down to the last ranked individual.

5. Results and Discussion

5.1 Citation Ranking

To develop a citation ranking for the study participants (A, B, C, D, E) and compare the ranking with peer evaluation data, all of the participants' citations (excluding self-citations) were compiled from the 350 scholarly works examined manually. Table 4 shows the frequency distribution of these citations ($n = 378$) and illustrates that citation counts are highly skewed: for example, participant A's citation count outnumbers that of participants D, C, E, and B by a ratio of 2.6, 3.1, 5.6, and 5.7 to 1, respectively. Recall that the study population consists of both associate (A and E) and full professors (B, C, D) and hence it might be possible that ranking by the total number of citations (straight count) is misleading because, all things being

⁹ Based on an investigation of their research and professional backgrounds, none of these 16 new individuals met our selection criteria.

TABLE 4. Participants' rankings by straight and adjusted citation counts.

Participant	Rank	Total no. of citations	Average no. of citations per year	Cumulative citation scores of all cited papers (no. of citations of a paper/age of paper)
A	1	183	12.2	34.6
D	2	71	6.5	18.5
C	3	59	5.9	13.4
E	4	33	4.7	10.8
B	5	32	2.3	6.1

Figures in parentheses refer to changes in ratios with respect to the ratios of total no. of citations.

equal, the longer a person has been conducting research in the field, the more citations he or she is likely to accumulate. To control for this possibility (or at least minimize its effect), data were normalized by the length of time each participant has been publishing on the Kurds. This was done by dividing the total number of citations by the number of years since publishing the first work on the Kurds (excluding dissertations). Accordingly, the average number of citations (adjusted count) compiled for each participant was calculated as follows: the total number of citations to all of A's publications was divided by 15 (years), B's citations by 14, C's by 10, D's by 11, and E's by 7.

In terms of ranking, the results remained the same. However, in terms of citations received per year (adjusted count), the ratios or the gaps between the participants changed. For example, the difference between participant A, on the one hand, and participants D and C, on the other hand, decreased by 27 and 32%, respectively. The difference between participants B and E changed from almost an equal number of citations (32–33) to more than double for participant E (4.7 citations/yr compared to 2.3/yr, respectively). One last example about the limitations of relying upon nonnormalized counts in ranking individuals can be observed in the case between participants A and E. While participant A outnumbered E by a ratio of 5.6 to 1 when straight citation counts are considered, the ratio drops to 2.6 to 1 when adjusted counts are compared. This is approximately a 54% decrease in difference. These results suggest that only normalized citation counts (taking into account time since publishing first paper after dissertation) should be used in ranking senior scholars, especially when current researchers are ranked and compared to one another. In other words, controlling for time should be applied to cases where researchers are still publishing on the topic but not to researchers who have ceased such an activity.

To test for the validity of the method applied above in normalizing citation counts, another method was investigated. Rather than dividing the total number of citations by the number of years since publishing their first work on the Kurds, the number of citations to each paper published by the participants was divided by the age of the respective paper. Citation scores for all cited papers were analyzed. Results obtained by this method were very similar to those found through normalization based on number of years since publishing first paper on the Kurds (see Table 4). This

finding suggests that results obtained from citation adjustment by time since publishing first paper (time-in-research, hereafter) is highly reliable.

Having established citation ranking for the participants, we next examine whether this ranking is similar to the three peer evaluation data sets collected in this study. If it is, then an argument can be made that normalized citation counts (taking into account time-in-research) serve as a valid indicator of research performance and hence is a valid measure to assess the quality and impact of senior faculty research. If not, we need to examine the differences and discuss the implications of this finding.

5.2 Citation Content

To generate different sets of peer evaluation data and compare them with citation data, three different types of sources are utilized: citation content, book reviews, and peer ranking. Starting with citation content, Table 5 illustrates the results from the content analysis of citations obtained by the study participants. The table shows that of the 378 citations analyzed, only five were negative in character. Given the ideological differences among the people who write on the Kurds' political and historical discourses and the emphasis that citation critics put on citation attitude, a much larger number of negative citations was expected. This was not the case here even though the study participants themselves represent three opposing viewpoints: independent (i.e., unbiased), pro-Kurdish, and anti-Kurdish, a classification derived from the book reviews examined for the present study. Whether this (i.e., the rarity of negative citing behavior) is a research norm among Kurdologists or whoever uses Kurdish literature is yet to be investigated. It should be mentioned though that earlier studies did find similar patterns. For example, MacRoberts and MacRoberts examined references in articles in the behavioral sciences and found that: "Generally speaking, criticism—and notably 'hostile criticism' as it is called in the literary circles—is rare" (1984, p. 91). They mention that criticism is avoided or disguised by three means: praise of the work criticized on grounds other than correctness, perfunctory citation, or avoidance of confronting any important living author who holds the view criticized. Possible reasons for this include fear of offending a friend or superior, negative reactions of reviewers to critical statements, and editorial attempts to

TABLE 5. Citation attitude data.

Respondent	Rank	Criteria							Total no. of positive and neutral citations*
		Positive and neutral				Negative			
		Paying homage	Recommending other readings	Supporting a personal idea	Value-free reference attitude	Criticizing a work	Disputing or disclaiming a work	Total no. of citations	
A	1	14 (7.7)	91 (49.7)	19 (10.4)	59 (32.2)			183	183
D	2	3 (4.2)	28 (39.4)	7 (9.9)	32 (45.1)	1 (1.4)		71	70
C	3	1 (1.7)	28 (47.5)	1 (1.7)	27 (45.8)	1 (1.7)	1 (1.7)	59	57
E	4	4 (12.1)	18 (54.5)	1 (3.0)	9 (27.3)		1 (3.0)	33	32
B	5	1 (3.1)	18 (56.3)	3 (9.4)	9 (28.1)		1 (3.1)	32	31
	Mean	5.8	49.5	6.9	35.7	0.6	1.6		

* Total positive score = (c1 + c2 + c3 + c4).

Figures in parentheses refer to the proportion in % of the citation(s) in that criterion with respect to the total number of citations.

moderate controversy. These patterns are the very opposite of what is found in book reviews examined in this study, as discussed below.

Having discovered that, i.e., having found very few negative citations, it was then believed that general citation behavior (positive vs. negative) does not influence citation ranking. We, therefore, focused on manipulating the citation scores of the positive and neutral criteria (paying homage, recommending other readings, supporting a personal idea, and value-free reference attitude) and the negative criteria (criticizing a work and disputing or disclaiming a work). This was done to examine whether a certain weighting scheme for combining the citations would generate different results for the participants or in ranking them as a whole. Various weighting schemes were calculated, however, all ended up with similar results. For example, the simplest scheme was to treat all positive and negative citations equally; that is, assigning one point (+1) for each positive citation and (−1) for each negative one. Because the number of negative citations was already found to be negligible, this scheme resulted in the same ranking as straight citation counts: 183, 70, 57, 32, and 31 for A, D, C, E, and B compared to 183, 71, 59, 33, and 32, respectively, in straight citation counts. Other weighting schemes based on the importance of the criteria were also calculated.¹⁰ However, none resulted in ranking changes or in any significant differences in scores among the participants.

The results above imply that, in general, it may be unnecessary to collect and analyze citation content; citation counts when accurately and thoroughly collected appear to provide the same information and require fewer resources to collect and analyze. While this may be true, a closer look at

the data also suggests that citation content may give some specific and important insights into the quality of research of some scholars, especially among those who rank in the middle. For example, of all C's citations, only 1.7% of them are of "paying homage" type whereas in the case of E, they are 12.1%. In addition, almost half of C's citations (45.8%) are of "value-free" type while E's are only 27.3% (see Table 5). While "paying homage" citations are clear indications of quality, "value-free" citations may or may not be so. Further analysis and research is needed to investigate this issue. In summary, it might be argued that citation content analysis is case-dependent and hence may or may not provide helpful or distinguishing insights onto the research performance of senior scholars.

5.3 Book Reviews

As mentioned earlier, the evaluative comments originating from the 44 book reviews examined in this study were classified according to 7 broad categories. Within each of these categories, comments were then grouped as positive or negative. Table 6 illustrates the distribution of these comments. Since the number of reviews for each book was not the same, adjusted scores for book review data were calculated by dividing the total number of positive comments compiled for each book by its respective number of reviews.¹¹ As in the case of citations, results from book reviews rank participant A at the top followed by participant D. Data in Table 6 further suggests that the books published by participants A and D have received more positive reviews in comparison with others and, therefore, are superior in the quality of the research they produce than others. The

¹⁰ For example, $(c1 + c2 + c3) \times 2 + c4$ and $(c1 \times 4 + c2 \times 3 + c3 \times 2 + c4 \times 1)$.

¹¹ Negative comments were not included because their number was found to be negligible.

TABLE 6. Data from book reviews.

Participant	Rank	Categories							Total score	Number of book reviews	Adjusted score (total score/no. of reviews)
		Originality and innovativeness	Analysis	Objectivity	Research	Sources	Readability	General attributes			
A	1	20	18 (21-3)	2	8	14	1	20	+83	12	+6.92
D	2	13	9 (14-5)	1	(1-1)	(8-3)	(1-1)	10	+38	10	+3.80
E	3	3	2 (5-3)		(1-1)	1		3	+9	3	+3.00
B	4	4	-6 (11-17)	1 (3-2)	-3	(4-5)	(1-1)	3 (6-3)	-2	9	-0.22
C	5	0 (3-3)	-8 (3-11)	2	-2	(4-4)	-1	-1 (4-5)	-10	10	-1.00

Figures in parentheses refer to the number of positive comments minus negative comments.

high quality of the books both A and D published is not only illustrated by the scores they obtain, but also by the relatively broad distribution of positive points across most categories. One more thing is noteworthy about the data illustrated in Table 6, namely, the sharp decline of participant C to the bottom of the list with a negative total score.

5.4 Peer Ranking

The last type of peer evaluation data collected was peer ranking. As mentioned earlier, peer ranking in this study was based on the opinions of 17 Kurdologists. Of these 17, only three were from the study participants: A, B, and D. This means that participants C and E had the advantage of the chance of being ranked by A, B, and D but not the other way round. To examine whether this has any impact on peer ranking, points were added to participants A, B, and D as follows: If any of A, B, or D has ranked C or E or both, he or she is given the same number of points. Results are illustrated in Table 7. Ranking was not affected at all, but scores did change slightly: participant A's score increased, but the differences between E and D as well as B and C decreased.

Overall, it was found that peer ranking puts participant A at the top of the list, followed by E (seventh), D (eighth), C (25th), and B (38th) when all 63 ranked Kurdologists are taken into consideration, or first, second, third, fourth, and fifth, respectively, when ranking is limited to the study

participants. What is interesting about this order is its resemblance to the results of book review data. In other words, both peer-based research evaluation systems data, i.e., book reviews and peer ranking, used in this study resulted in similar rankings. First, participant A is top-ranked by his peers and is also ranked considerably ahead of the others. Similarly, in the book reviews, participant A is ranked at the top and well ahead of the other participants in terms of positive comments. Second, both participants D and E are ranked similarly. Third, both participants B and C are once again found at the bottom of the list with a large gap between them and participants A, D, and E.

5.5 Correlation Between Citation Ranking and Peer Evaluation Data Sets, and the Validity of Citation Ranking in Evaluating the Research Performance of Senior Faculty Members

Using citation ranking, citation content, book review, and peer ranking data, we can now address our research questions: (1) To what degree does citation ranking correlate with data from citation content analysis, book reviews, and peer ranking? (2) Is citation ranking a valid evaluative indicator of research performance of senior faculty members?

Data from Tables 4, 5, 6, and 7—summarized in Table 8—show that, relatively speaking, there are three distinct groups of scholars among the study population: participant A with excellent research performance, participants D and E with good research performance, participants B and C with poor research performance. Such a classification is typically made by hiring, promotion, and tenure committees when making decisions or recommendations regarding applicants or candidates.

Recalling the assumption made above that peer evaluation is a valid and the most accepted indicator of scholarly research performance, it appears that normalized citation data (or average citations per year) performed as good, especially in evaluating those at the high and low ends of

TABLE 7. Peer ranking.

Participant	Peer ranking*	Peer ranking (scores)	Adjusted scores
A	1	293	306
E	7	174	174
D	8	139	155
C	25	80	80
B	38	48	59

* Overall ranking of the participants among the top 63 Kurdologists.

TABLE 8. Summary of rankings of participants according to data sets collected.

Respondent	Adjusted citation count (per year)	Positive and neutral citation behavior scores	Adjusted book reviews scores	Adjusted peer ranking scores
A	12.2	183	+6.92	306
D	6.5	70	+3.80	155
E	4.7	57	+3.00	174
C	5.9	32	-1.00	80
B	2.3	31	-0.22	59

the relative rankings—participants A, D, and B. For example, participant A is ranked at the top in terms of citation counts as well as in book reviews and peer ranking. Participant D is ranked second in terms of citations and book reviews and third in peer ranking. Participant B is ranked last in terms of citations, a performance very similar to what peer evaluation data sets revealed (see Figure 1). In the middle group, participant E ranks fourth in citations, third in book reviews and second in peer ranking, while participant C ranks third in citations, last in book reviews, and fourth in peer ranking.*

The conclusion reached above—that normalized citation data perform as well as peer judgments in evaluating the research performance of scholars at the high and low ends of the relative rankings—was further supported by analyzing correlations between citation ranking, book review data, and peer ranking. Results show that there is a statistically significant correlation between citation ranking and book review ranking (Spearman's $r = 1.000$), and between citation ranking and peer ranking (Spearman's $r = 1.000$) for participants A, D, and B—the two most highly ranked scholars and the lowest ranked scholar overall.

On the otherhand, results show that there is no statistically significant correlation between citation ranking and book review ranking (Spearman's $r = .700$), and between citation ranking and peer ranking (Spearman's $r = .700$), when data from all scholars are considered. Having found this, it is, therefore, worthwhile to investigate why citation ranking does not perform particularly well for mid-range performers. In particular, we need to investigate why C performs well in citations but not in book reviews and peer ranking, and why participant E performs well in book reviews and peer ranking but not as well in citations?

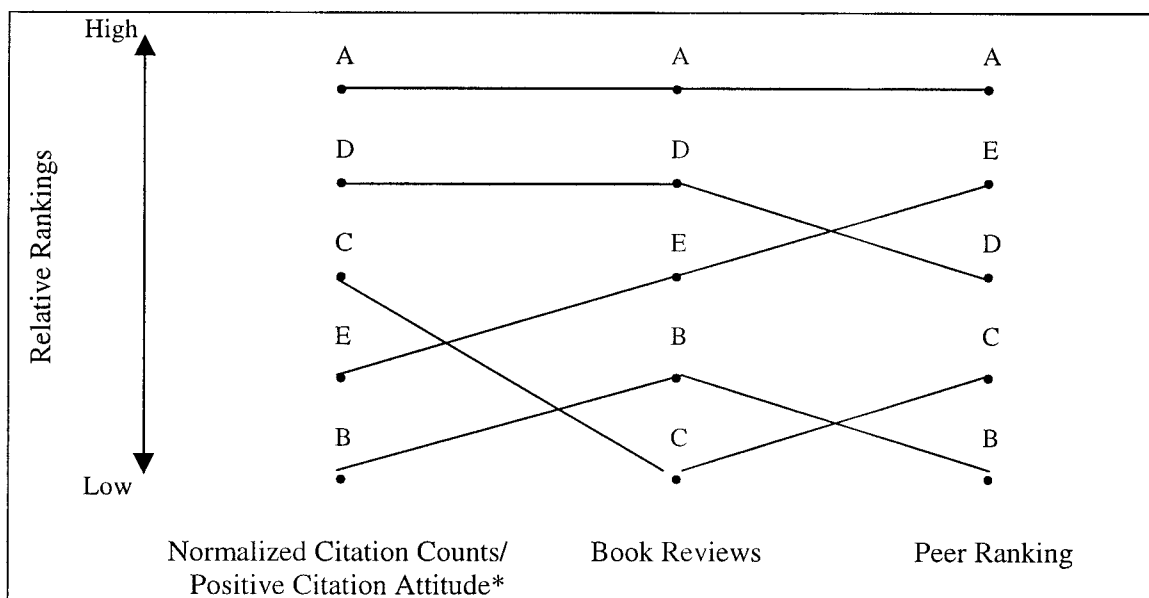
There could be many possible reasons for this. One could be attributed to the research focus of the participants. Given that the number of publications and researchers who publish on the political and historical discourses of the Kurdish question is much larger than in any other area in Kurdish studies (see Meho, 1997), it is possible that historians and political scientists have larger chances of being cited than other researchers. A closer look at the full list of publica-

tions of C and E reveal that all of C's works on the Kurds ($n = 24$) deal exclusively with the political and historical discourses of the Kurdish question, whereas only four out of 12 of E's works do; the remaining eight works are primarily communication- and education-related research. This suggests that participant C has a larger audience of researchers for citing the papers s/he produces than participant E. This, in turn, suggests that E's research is of high quality but not "mainstream" and C's work is "timely" and "mainstream" but not of high quality (as judged by peers). A closer look at why each of E and C is cited further validates this conjecture. Almost two-thirds (67%) of E's citations are explicitly quality-related—citations for paying homage to peers and those which recommend the study participants' publications for further reading—whereas only 49% of the citations for C are explicitly quality-related (see Table 5). Also worthwhile to mention is that the ten papers published by C between 1993 and 1998 have received only two citations in total (as of December 1998) while E's four published papers in the same period received six citations in total; that is, three-fold the number of citations for a smaller number of publications. Furthermore, of the 24 papers published by C through 1998, only 10 (or 41%) received citations and four (or 17%) received more than one citation. In the case of E, 10 (or 83%) of his or her 12 papers received citations and six (or 50%) received more than one.

Another reason which might explain why C performs well in citations but not in book reviews and peer ranking, and why participant E performs well in book reviews and peer ranking but not as well in citations could be attributed to the number of researchers who have access to C and E books. Meho and Haas (in preparation) found that social science Kurdologists,* in general, cite books much more than any other type of materials (46%). This was followed by newspaper articles and broadcast news (20%), journal articles (18%), and government publications (9%). This implies that books are among the most frequent sources of citations and thus books availability to a larger audience might largely enhance the chances for being cited. According to OCLC, participant C's books are available in many more libraries than E's book (429 compared to 47 libraries,

* Positive and neutral citation behavior ranking was not mentioned because rankings derived from citation counts and citation behavior were identical.

* Include faculty from the fields of anthropology, area studies, communication, education, geography, history, political science, sociology, and women studies.



*Normalized citation counts and positive citation attitude are merged together because they yielded identical ranking results.

FIG. 1. Comparative relative rankings for study population.

respectively). Assuming that all citations to their books come from users who have access to the books, these figures might explain why C is more cited than E.[†]

In summary, it appears that normalized citation ranking (taking into account time-in-research), citation content analysis, book reviews, and peer ranking perform similarly for higher-ranked and lower-ranked senior scholars. Additional evaluation methods and measures that take into account the context and content of research appear to be needed to effectively evaluate senior scholars whose performance ranks relatively in the middle. Citation content analysis may provide specific and important insights onto the research quality of scholars whose performance ranks relatively in the middle, however, further analysis and research is needed to investigate this issue. Finally, normalizing citation counts according to time-in-research proved to be better than raw citation counts.

Although we cannot claim that citations, book reviews, and peer ranking are measuring the same thing, all taken together, however, one may argue that they collectively serve as measures for high and low research performance. For example, as discussed earlier, not all citations are indicators of quality, yet more than 55% of them were clearly related to quality—citations for paying homage to peers and those which recommend the study participants' publications for further reading (see Table 5). On the other hand, book

reviews may appear to thoroughly evaluate only the items at hand. In reality, however, the reviews are evaluating a major portion of the study participants' publications. As stated above, books published by the participants and examined here are primarily compilations to several previously published articles over a long period of time. Finally, peer ranking is an indicator that tends to take a person's whole academic career into consideration rather than a particular time period or publication.

6. Toward Validity and Generalizability

To investigate the validity and generalizability of the results of our study, we compared citation and peer ranking data for three additional groups of senior Kurdologists:

- Group 1: Kurdologists who publish primarily in English and on the historical and political discourses of the Kurdish question ($n = 9$). These include our five study participants, two nonfaculty members, and two who published their books in 1997–1998.
- Group 2: French-publishing Kurdologists ($n = 6$). Fields of these Kurdologists include history, linguistics, literature, and politics. All have been publishing for at least 15 years.
- Group 3: Humanities Kurdologists ($n = 9$). Fields of these Kurdologists include linguistics, literature, and religion.

These groups were selected because they represent other major groups of contemporary Kurdologists. Citation and peer ranking data for members of these groups were com-

[†] Discussing factors that libraries take into consideration before buying a book and the market share or marketing skills of the publishers is beyond the scope of this paper, yet, one important aspect should be mentioned: according to Books in Print, E's book is seven times more expensive than any of C's books, something which might explain why E's book is available in far fewer libraries.

TABLE 9. Citation ranking and peer ranking scores for the three groups of Kurdologists.

ID	Citation score	Peer ranking score
Group I (history-politics)—Average citations per year		
A	16.5	293
I	10.7	185
D	10.0	139
C	8.2	80
II	7.0	60
E	6.6	174
III	2.9	51
B	2.8	48
IV	2.0	26
Group II (French-writing Kurdologists)—Total number of citations		
i	51	260
ii	47	93
iii	39	103
iv	30	134
v	25	87
vi	6	40
Group III (Humanities Kurdologists)—Total number of citations		
a	48	137
b	45	131
c	39	103
d	30	134
e	20	47
f	16	55
g	6	40
h	5	8

piled from the same peer ranking survey discussed previously and citations from SSCI, AHCI, and the 350 scholarly works examined initially for our original study participants. The average citations per year were computed for the mem-

bers of the first group only. The reason for this is because all of the members of this group were among the younger generation of Kurdologists, their publication start date was easily determined (from the mid-1980s to early 1990s), and because they are still publishing on the Kurds with five being from our original study participants. The average citations per year were not computed for the members of the other two groups primarily because they are of the older generation of Kurdologists and hence it was not plausible to normalize their citation performance according to publication start date—which ranges from the late 1950s to early 1970s. Therefore, whole citation counts were compiled for members of groups 2 and 3 (see Table 9).

To test our finding that citation ranking performs best for highly-ranked and lower-ranked senior scholars rather than middle performers, we calculated citation ranking for individuals in each of the three groups and compared their citation ranking to their respective peer ranking. We did not include a citation content analysis because the original analysis illustrated a 100% correlation between citation counts and citation content analysis. Furthermore, we did not conduct a book review content analysis because a few individuals in these groups had not published a book and still others did not have any reviews for their books. The data from the three groups showed patterns similar to our initial data. That is, the matching or the correlations between citation ranking and peer ranking among all groups consistently showed that citation ranking perform well in assessing the research quality of highly-ranked and lower-ranked senior scholars but not particularly well for midrange performers.

For example, Figure 2 illustrates that both of the top-cited scholars (A and I) and the three least cited ones (III, B, IV) of the Kurdologists who write on the historical and political discourses of the Kurds are ranked in the same order as judged by peers. Scholars who are relatively ranked

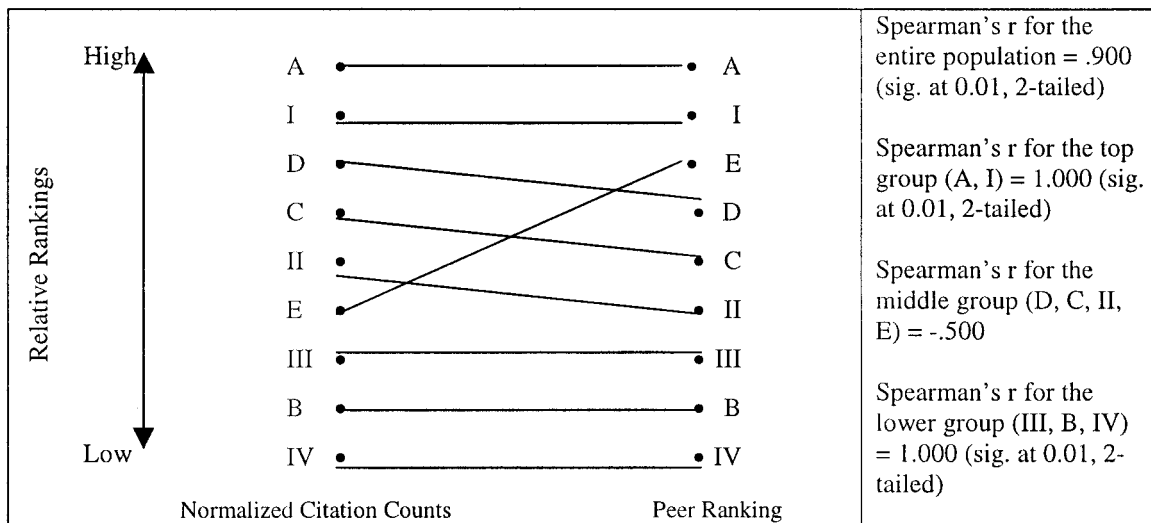


FIG. 2. Comparative relative rankings for Group 1, history-politics senior Kurdologists ($n = 9$).

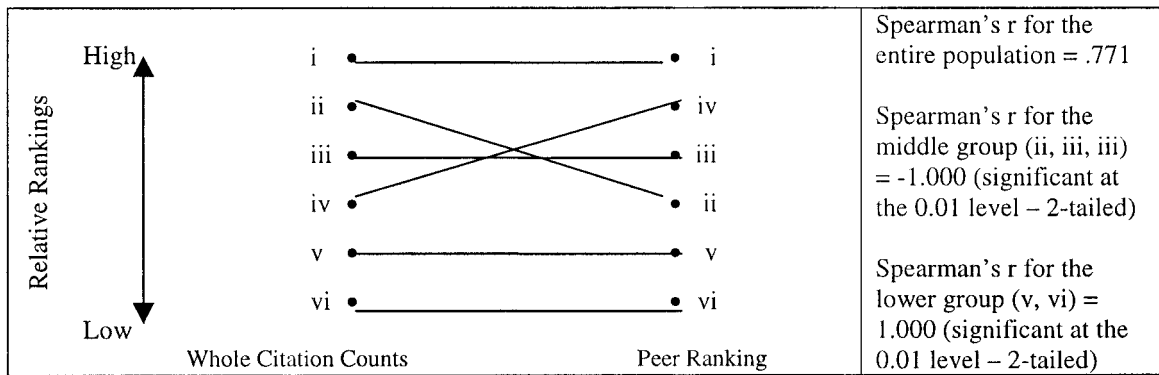


FIG. 3. Comparative relative rankings for Group 2, French-writing senior Kurdologists ($n = 6$).

in the middle in terms of citations are also ranked so by their peers, however, in different orders. While D, C, II, and E are respectively ranked third, fourth, fifth, and sixth in terms of citations, in peer ranking, the order was fourth, fifth, sixth, and third, respectively. The greatest difference recorded between the two evaluation methods was for scholar E, who ranked sixth in citations but averaged third according to peers. In faculty funding, hiring, promotion, and tenure decisions, this is a big difference. Usually in these committees, only a small number of applicants make it to the final stages. Therefore, additional evaluation methods and measures appear to be needed to effectively evaluate the research performance of these middle performers.

Consistent patterns were found among the members of the other two groups—French-writing and humanities Kurdologists. That is, similar to members of Group 1, top-cited and least-cited scholars in Group 2 and Group 3 are ranked in the same order as judged by peers. Also, middle performers in these two groups are ranked differently when citation and peer ranking are compared (see Figures 3 and 4). These results further support the validity of the proposition that citation ranking (or counts) perform best for high-ranked

and lower-ranked senior scholars rather than middle performers. The results also indicate that our findings are generalizable to other populations.

7. Conclusion

The purpose of this study was to analyze the relationship between citation ranking and peer evaluation in assessing senior faculty research performance. Two main questions were investigated: (a) To what degree does citation ranking correlate with data from citation content analysis, book reviews, and peer ranking? (b) Is citation ranking a valid evaluative indicator of research performance of senior faculty members? Results presented evidence that adjusted citation ranking, taking into account time since publishing first scholarly paper after dissertation, can be an effective evaluation measure of research performance for current senior faculty members, especially for top-ranked and low-ranked individuals. This was demonstrated by the high correlation and ranking match between citations and book reviews, and citations and peer ranking. This study provides an example of how different evaluative data sets (in this

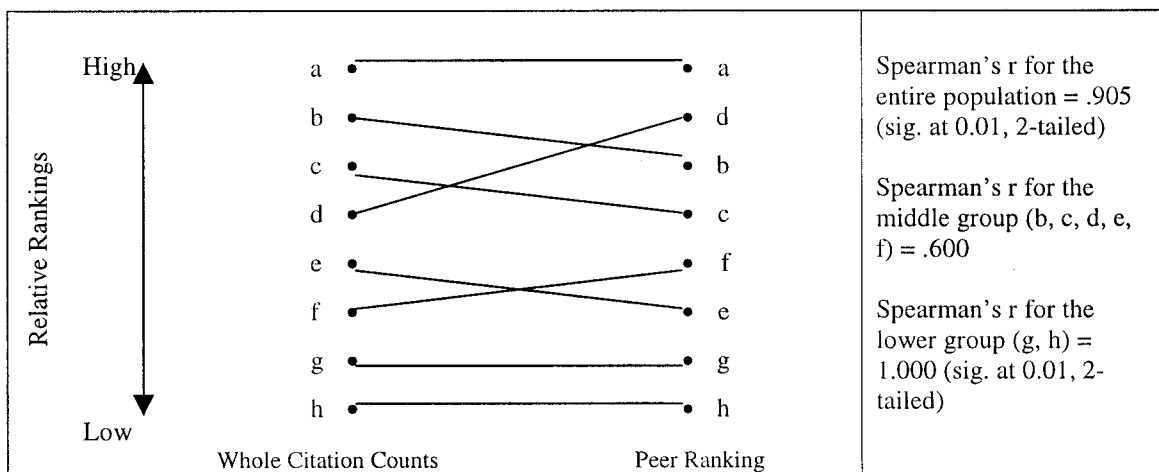


FIG. 4. Comparative relative rankings for Group 3, humanities senior Kurdologists ($n = 9$).

case, citations, book reviews, and peer ranking) can be used in real life situations when hiring, promotion, and tenure committees are faced with the task of evaluating senior candidates.

The results presented in this study illustrate the validity of using citation data to evaluate the research performance of individual senior scholars. Results show that normalized citation counts and citation content analysis data yield identical rankings. Results also show that citation ranking, citation content analysis, book reviews, and peer ranking perform similarly for higher-ranked and lower-ranked senior scholars. Additional evaluation methods and measures that take into account the context and content of research appear to be needed to effectively evaluate senior scholars whose performance ranks relatively in the middle. These findings were further corroborated by analyzing citation and peer ranking data for three additional groups of senior scholars suggesting that the results are valid and are generalizable to other areas of scholarship.

Although more studies, including studies with a larger sample size and researchers in other disciplines, are needed to further verify the generalizability of these results, at present we should at least feel more confident about the relative success of the citation ranking method in research evaluation systems, especially when senior scholars in the context of smaller research specialty areas are compared one to another. Citation ranking can provide a valid assessment tool for comparative research evaluation and hence should be used as a relative measure in confirming or questioning peer-based research evaluation assessments of senior faculty to arrive at a better judgment. Finally, it is hoped that this study will be replicated to reach a better understanding of the role of citations in research evaluation systems.

Acknowledgments

We would like to thank Bob Losee for his insightful comments. We would also like to thank our study participants, those who responded to our peer ranking survey, and JASIS reviewers.

References

Allen, B. (1989). Propositional analysis: A tool for library and information science research. *Library and Information Science Research*, 11(3), 235–246.

Allen, B., & Reser, D. (1990). Content analysis in library and information science research. *Library and Information Science Research*, 12(3), 251–262.

Arts & Humanities Citation Index (1980–1998). Philadelphia: Institute for Scientific Information.

Baird, L.M., & Oppenheim, C. (1994). Do citations matter? *Journal of Information Science*, 20(1), 2–15.

Berelson, B. (1952). *Content analysis in communication research*. Glencoe, IL: Free Press.

Bilhartz, T.D. (1984). In 500 words or less: Academic book reviewing in American history. *The History Teacher*, 17, 525–536.

Book Review Digest (1905–1998). New York, NY: The H.W. Wilson Company.

Book Review Index (1965–1998). Detroit, MI: Gale Research Co.

Books in Print (1948–1998). New York, NY: R.R. Bowker Co.

Bornstein, R.F. (1991). Manuscript review in psychology: psychometrics, demand characteristics, and an alternative model. *The Journal of Mind & Behavior*, 12, 429–468.

Busha, C.H., & Harter, S.P. (1980). *Research Methods in Librarianship: Techniques and Interpretation*. New York, NY: Academic Press.

Casey, J.B. (1985). *Assessment of Quality in Book Selection: An Evaluation of the Effectiveness of Opinions Rendered by Peer Reviews in American History Journals*. PhD Thesis, Case Western Reserve University, Cleveland, OH.

Champion, D.J., & Morris, M.F. (1973). A content analysis of book reviews in the *AJS*, *ASR*, and *Social Forces*. *American Sociological Review*, 78, 1256–1265.

Chubin, D.E., & Hackett, E.J. (1990). *Peerless science: Peer review and U.S. science policy*. Albany, NY: State University of New York Press.

Clark, K.E. (1957). *America's psychologists: A survey of a growing profession*. Washington, DC: American Psychological Association.

Cole, S., & Cole, J.R. (1967). Scientific output and recognition. *American Sociological Review*, 32, 377–390.

Cole, S., & Cole, J.R. (1968). Visibility and the structural bases of awareness of scientific research. *American Sociological Review*, 33, 397–413.

Cronin, B., & Overfelt, K. (1994). Citation-based auditing of academic performance. *Journal of the American Society for Information Science*, 45(2), 61–72.

Garfield, E. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*. New York, NY: John Wiley & Sons, Inc.

Garfield, E. (1983). How to use citation analysis for faculty evaluations, and when is it relevant (Pts. 1 & 2). *Current Contents*, 44, 5–13; 45, 5–14.

Gilbert, G.N. (1978). Measuring the growth of science: A review of indicators of scientific growth. *Scientometrics*, 1, 9–34.

Glanzel, W. (1996). The needs for standards in bibliometric research and technology. *Scientometrics*, 35, 167–176.

Glenn, N.D. (1978). On the misuse of book reviews. *Contemporary Sociology*, 7, 254–255.

Hicks, C.E., Rush, J.E., & Strong, S.M. (1985). Content analysis. In E.D. Dym (Ed.), *Subject and information analysis* (pp. 57–109). New York, NY: Marcel Dekker.

Hirsch, W., Kulley, A.M., & Efron, R.T. (1974). The gatekeeping process in scientific communication: Norms, practices and content of book reviews in professional journals. Working Paper Number 81, Purdue University, Institute for the Study of Social Change, West Lafayette, IN.

Historical Abstracts (1955–1998). Santa Barbara, CA: American Bibliographical Center of ABC-Clio.

Holsti, O.R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley Pub. Co.

Index Islamicus (1906–1998). East Grinstead, West Sussex, UK: Bowker-Saur.

Ingram, H.M., & Mills, P.B. (1989). Reviewing the book reviews. *PS: Political Science and Politics*, 22, 627–634.

International Political Science Abstracts (1951–1998). Oxford: Basil Blackwell; Paris: Presses Universitaires de France.

International Society Kurdistan (1968). *ISK's Kurdish bibliography*. S. van Rooy, & K. Tamboer (Eds.), Amsterdam.

Irvine, J., & Martin, B.R. (1983). Assessing basic research: The case of the Isaac-Newton telescope. *Social Studies of Science*, 13, 49–86.

King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science*, 13, 261–276.

Koenig, M.E.D. (1982). Determinants of expert judgment of research performance. *Scientometrics*, 4, 361–378.

Koenig, M.E.D. (1983). Bibliometric indicators versus expert opinion in assessing research performance. *Journal of the American Society for Information Science*, 34, 136–145.

- Kostoff, R.N. (1996). Performance measures for government-sponsored research: overview and background. *Scientometrics*, 36, 281–292.
- Lawani, S.M., & 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*, 34, 59–66.
- Lindholm-Romantschuk, Y. (1998). *Scholarly book reviewing in the social sciences and humanities: The flow of ideas within and among disciplines*. Westport, CT: Greenwood Press.
- Liu, M. (1993). The complexities of citation practice: A review of citation studies. *Journal of Documentation*, 49, 370–408.
- MacRoberts, M.H., & MacRoberts, B.R. (1984). The negational references: Or the art of dissembling. *Social Studies of Science*, 14, 91–94.
- MacRoberts, M.H., & MacRoberts, B.R. (1986). Quantitative measures of communication in science: A study of the formal level. *Social Studies of Science*, 16, 151–187.
- MacRoberts, M.H., & MacRoberts, B.R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, 40, 342–349.
- MacRoberts, M.H., & MacRoberts, B.R. (1996). Problems of citation analysis. *Scientometrics*, 36, 435–444.
- Martin, B.R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, 36, 343–362.
- Meho, L.I. (1997). *The Kurds and Kurdistan: A selective and annotated bibliography*. Westport, CT: Greenwood Press.
- Meho, L.I., & Haas, S.W. (in preparation). The importance, use, and nonuse of government publications by social scientists studying stateless nations.
- Moxley, J.M. (1992). *Publish, don't perish: The scholar's guide to academic writing and publishing*. Westport, CT: Greenwood Press.
- Mulkay, M.J. (1974). Methodology in the sociology of science. *Sociology of Science*, 13, 107–119.
- Myers, R.C. (1970). Journal citations and scientific eminence in contemporary psychology. *American Psychologist*, 25, 1041–1048.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ: Computer Horizons.
- Narin, F., & Hamilton, K.S. (1996). Bibliometric performance measures. *Scientometrics*, 36, 293–310.
- Natowitz, A., & Carlo, P.W. (1997). Evaluating review content for book selection: An analysis of American history reviews in *Choice*, *American Historical Review*, and *Journal of American History*. *College & Research Libraries*, 58, 323–336.
- PAIS International (1915–1998). New York, NY: Public Affairs Information Service.
- Peritz, B.C. (1983). A classification of citation roles for the social sciences and related fields. *Scientometrics*, 5, 303–312.
- Political Science Abstracts (1975–1998). New York, NY: IFI/Plenum.
- Riley, L.E., & Spreitzer, E.A. (1970). Book reviewing in the social sciences. *The American Sociologist*, 5, 358–363.
- Schwartz, C.A. (1989). Book selection, collection development, and bounded rationality. *College & Research Libraries*, 50, 328–343.
- Seglen, P.O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43, 628–638.
- Seglen, P.O. (1998). Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthopaedica Scandinavica*, 69, 224–229.
- Smith, L.C. (1981). Citation analysis. *Library Trends*, 30, 83–106.
- Snizek, W.E., & Furham, E.R. (1979). Some factors affecting the evaluative content of book reviews in sociology. *The American Sociologist*, 14, 108–14.
- Social Sciences Citation Index (1972–1998). Philadelphia: Institute for Scientific Information.
- Sociological Abstracts (1963–1998). San Diego: Sociological Abstracts, Inc.
- Spiegel-Rosing, I. (1977). Science studies: Bibliometric and content analysis. *Social Studies of Science*, 7, 97–113.
- Stieg Dalton, M. (1995). Refereeing of scholarly works for primary publishing. *Annual Review of Information Science and Technology*, 30, 213–250.
- Taylor, W.R. (1967). The influence of professional status differences upon book reviewing in sociology. Ph.D. Thesis. Purdue University, West Lafayette, IN.
- Thomas, P.R., & Watkins, D.S. (1998). Institutional research rankings via bibliometric analysis and direct peer review: a comparative case study with policy implications. *Scientometrics*, 41, 335–355.
- van Raan, A.F.J. (1996). Advanced bibliometric methods as quantitative core of peer-review based evaluation and foresight exercises. *Scientometrics*, 36, 397–420.
- van Raan, A.F.J. (1997). *Scientometrics: State of the art*. *Scientometrics*, 38, 205–218.
- Virgo, J.A. (1977). A statistical procedure for evaluating the importance of scientific papers. *Library Quarterly*, 47(4), 415–430.
- White, H.D. (1990). Author co-citation analysis: Overview and defense. In C.L. Borgman (Ed.), *Scholarly communication and bibliometrics* (pp. 84–106). Newbury Park, CA: Sage.
- WorldCat [computer file, 1200–1999]. Dublin, OH: OCLC.