CHAPTER – 2
REVIEW OF LITERATURE
2. REVIEW OF LITERATURE

The present chapter, review of literature, focuses on theories and studies that have formed on the basis of citation analyses. In broad concept, the term citation is used synonymously with the term bibliographic reference. Citation analysis is performed to represent the analysis of bibliographic references, which form part of the apparatus of scholarly communication. Citation analysis and various citations based indexing and abstracting services are different from each other. Martyn (1975) stated that “citation in the primary literature expressly states a connection between two documents, one which cites and the other which is cited, whereas citation in other listings does not usually imply any connection between documents other than that effected by the indexing machinery.”

Citation based analysis was started by E. Garfield and the idea of an impact factor was introduced in Science in 1955 (Garfield, 1955). With support from the National Institutes of Health, the experimental Genetics Citation Index was published, and that led to the 1961 publication of the Science Citation Index (Garfield & Sher, 1963). An article titled “The Norms of Citation Behaviour” was introduced for the first explicit account of citing as normative behavior by Kaplan (1965). According to him, the major function of footnoting practice was the reaffirmation of the underlying norms of scientific behavior.

Evaluative citation analysis was started to point to evidence of the correlation of citation counts with other measures of the quality of documents or their authors’ productivity peer ratings, and awards of grants and prizes (Bayer & Folger, 1966). Knowledge about citing behavior and the symbolic characteristics of citations is essential in order to determine whether it makes sense to use citation analysis in
various areas of application. As Zunde (1971) noted, citation analysis has three main applications: (i) qualitative and quantitative evaluation of scientists, publications, and scientific institutions; (ii) modelling of the historical development of science and technology; (iii) information search and retrieval. The introduction of two special citation-analytical techniques has covered the way for a significant application of Knowledge Organization based on bibliographic coupling (Kessler, 1963) and co-citation analysis (Marshakova, 1973).

Cawkell (1974) claimed that the deductions that can be made from a citation network without knowledge of its subject content. There was a need for a theory of citing that could explain why authors cite the way they do. A first contributor was Mulkay (1974) who argued that there had been no clear demonstration of the way in which citations reflect the process of scientific influence. He concluded that “in fact we know very little about who cites whom in science and why”. Studies of citations appeared in abstracting and indexing services, in subject bibliographies were first noted by Martyn (1975) who argued that “citation in the primary literature expressly states a connection between two documents one which cites and the other which is cited, whereas citation in other listings does not usually imply any connection between documents other than that effected by the indexing machinery.”

To find the motives of citations, Moravcsik and Murugesan (1975) conducted a study using 706 references in 30 articles on high-energy physics. Their study revealed, among other things, that 41% of the references were nonessential and 14% were negative. Smith (1981) listed a general set of basic assumptions for citation analysis: (i) citation of a document implies use of that document by the citing author; (ii) citation of a document (author, journal, etc.) reflects the merit (quality, significance, impact) of that document (author, journal, etc.); (iii) citations are made
to the best possible works; (iv) a cited document is related in content to the citing
document and (v) all citations are equal.

A research investigated the structure and the interrelations of the library and
information science journal literature as a study on the flow of information in library
and information science. It has been proved that there are the distinct differences
between the citations obtained from the journals in the field of library science and
those of information science (Saito, 1980). In the library science, citations from all of
the articles in Library Journal, Library Resources and Technical Services, Special
Libraries, and College and Research Libraries are preferred and in the information
science, JASIS and Journal of Documentation are examined. The main database is
Social Sciences Citation Index (1970-1980). There are some differences in basic
literatures of two disciplines; (i) articles in the information science cite basic
literatures more regularly, (ii) most of the basic literature in the information science
was published after 1940, (iii) there are few duplicated books in two disciplines
(Ueda, Miwa & Nakayama 1983).

The co-citation analysis assumes that the more often two documents are cited
together, the closer the relationship between them. Saito (1984) discussed the
intellectual structure of a scientific specialty using Kuhn's paradigm theory. The
specialty is considered a set of paradigm-sharing communities. In this viewpoint the
co-cited authors and the co-cited papers are regarded as the research frontiers in the
specialty and the work undertaken by the member experts.

Various theoretical models were analyzed for motives of citation behaviour by
Brooks (1985). His survey found that expressiveness was the most common
purpose for citing and that only 2% of the references were refusal. A study of the
incidence and variety of bibliographic errors found that almost a quarter of the references contained at least one mistake and 8 percent of these were judged serious enough to prevent retrieval of the article (De Lacey, Record & Wade, 1985). Social constructivist theory of citing explains that successful scientists are those who most skillfully manage to persuade others that they are not just being persuaded, that no mediations intercede between what is said and the truth (Latour & Woolgar, 1986). According to this view, when authors cite, they are positioning earlier documents in such a way as to convince readers of the goodness of their claims.

An important feature of citations is that each reference is an inscription describing a certain text by a standardized code. Although different publication manuals give different codes, many publishers of books and journals follow their own standards (Latour & Woolgar, 1986). A study was conducted to assess the dissemination and impact of the projects in the professional literature. The citations tended to be gathered among a small number of library-related serials. A small number of funded projects accounted for a large number of the citations. The most cited projects cost only one-fifth as much as the most expensive studies, yet were cited nearly five times as often (Altman & Antieau, 1988).

In 1998, a paper was published on the title “Theories of Citation?” by taking the whole issues of the journal Scientometrics. The author concluded that although a variety of contexts for citation analysis had been proposed, a comprehensive theory had not been formulated (Leydesdorff, 1998).

To find out a formula or objective guideline to determine the correct allocation ratio for serials vs. monographs more than 50 studies over the past 40 years were performed. Devin and Kellogg (1990) used citation analysis to determine the actual
percentage of serials vs. monographs used by researchers in each field. A citation analysis was conducted to identify factors associated with subjective rankings of a journal's value in promotion and tenure decisions. Prestige rankings from a 1982 survey of ARL directors and library school deans were correlated with nine citation measures: Total citation count, impact factor, immediacy index, references per paper, Price's Index, self-citation rate, popularity factor, citation factor, and consumption factor, with and without controlling for journal orientation, age, circulation, and index coverage (Kim, 1991).

To study the accuracy of citations in library literature, a research was carried out by examining 1,094 references from 131 articles. In 193 references, 223 errors were detected. A review of citations at manuscript stage was also carried out for one of the journals. The study highlighted a need for greater awareness among LIS professionals of keeping their citations error free, and suggested other aspects of citations (Pandit, 1993). To understand the recency of citations and to determine the pattern of literature usage by Indian agricultural scientists a study was conducted by Munshi and Vashishth (1993). The cost and effort of the Research Assessment Exercise may not be justified when a simpler and cheaper alternative, namely a citation counting exercise, could be undertaken (Oppenheim, 1995).

A study for analyzing self-citation in the library and information science literature was conducted by Dimitroff and Arlitsch. Articles that were reports of research that were written by a faculty member, that addressed a theoretical topic, or that had multiple authors were all more likely to have to higher self-citation rates (Dimitroff & Arlitsch, 1995). Oppenheim (1995) showed that there is a statistically significant correlation between the numbers of citations received by a department in total and the Research Assessment Exercise rating. This provides further independent
support for the validity of citation counting, even when using just the first authors as a search tool for cited references.

Using Social SciSearch on DIALOG to analyze citations to twenty-four library science journals over a twenty-year period Meyer and Spencer (1996) identified that the non-library science fields that cited articles published in the library journals included in their study by using the journal subject categories on DIALOG. Although citations from other fields are higher than previous studies indicated, comparison with other fields in the social sciences shows that library science is not commanding citations at the level of the more developed fields. To investigate one aspect of library collection use, a study was completed taking citations from theses and dissertations. The Kendall coefficient of rank correlation tests the degree of association between journals most heavily cited by graduate students and those titles most heavily cited in faculty publications. Positive associations are confirmed in three data sets (Zipp, 1996).

Loughner (1996) used citation analysis as helpful tool for evaluating academic library usage. This study revealed how to use the Science Citation Index CD-ROM product and a personal computer to generate useful reports utilizing a much larger base of citations (Crotteau, 1997).

To investigate how much e-sources have been used in formal scholarly communication, a case study was completed in the area of library and information science during the period 1994 to 1996. Four citation-based indicators were proposed and used in this study for the impact measurement. Compared with the impact of print sources, the impact of e-sources on formal scholarly communication in LIS was small,
as measured by e-sources cited, and does not increase significantly by year even though there is observable growth of the impact across the years (Zhang, 1998).

In an attempt to create a new tool, Kuyper (1999) performed a study to aid librarians in choosing music journals. All the citations from music dissertation bibliographies submitted in 1993 from across the United States were gathered and analyzed and listed. Core lists of journals were developed and then compared to the said lists. The journal lists from a national study differed from those derived from the study of journals used at a single institution. New academic programmes in environmental science prompted a citation analysis to characterize the citation patterns of the interdisciplinary field of environmental and human health as compared with other disciplines. The results showed that the average age of citations was 10.5 years for journals and 9.4 years for books. On average, journals were cited 67% of the time while books were cited 17% of the time (Johnson, 1999).

A citation based study described the development and application of visualization techniques for users to access and explore information in a digital library effectively and intuitively. It also showed author co-citation patterns through a number of author co-citation maps (Chen, 1999).

An analytical study explored the citations age distribution and the difference between them. The study illustrated that the age distribution for the mean of all the journals is an exponentially decaying curve and the citation age distributions show a sharp initial rise from age one to three or four years and then fall off in a sort of exponential decay (Tsay, 1999). An investigation showed the potential value of bibliometric analysis of student research paper. In this study Hovde (2000) suggested
that student paper bibliographies provide a flexible, non-invasive, time-efficient assessment forum for the documentation of student library use.

One study explored the advantages and disadvantages in case of using JCR citation data for journal management. This study revealed the impact of journal self-citation on JCR rankings of library and information science (LIS) and genetics journals. The impact factor and total citations received were recalculated and removed self-citations. The study concluded that librarians can use JCR data without correcting for journal self-citation, although self-citations do exert a major effect on the rankings for a small number of journals (Nisonger, 2000). Vreuland (2000) suggested that the use of citation analysis is an appropriate objective standard for measuring and evaluating law library web sites.

A research was carried out by Davis (2002) through citation analysis. The result showed that the journals from society and associations received the highest number of citations and were priced considerably lower than commercial journals. However, most of these studies suffer from a number of problems. Their biggest problem is that they have often focused narrowly on the opposite extremes of the citation distributions. Only a few studies have dealt with entire citation distributions. Moreover, Bornstein’s hypothesis about a J-shaped relationship between research quality and citation counts has recently received empirical confirmation (Nicolaisen, 2002).

Jacso (2003) showed that Yahoo has abandoned subscribers of its Research Document Collection. He showed that there are two databases available that are equally good- one is the BizTech Library and the other is the Energy Citations Database. A study was conducted to investigate the relationship between the journal
"impact factor" and the "immediacy index" It was found that there are highly positive correlations between them (Yue, Wilson & Rousseau, 2004).

Swygart-Hobaugh (2004) examined that how the social sciences' debate between qualitative and quantitative methods is reflected in the citation patterns of sociology journal articles. The results of a quantitative citation analysis showed puzzling differences concerning female and male authors' citation practice (Hakanson, 2005). To study the citation checking of advanced undergraduate research papers as a method for evaluating library collections a study was conducted with a sample 101 honours theses. It showed that the proportion of journal citations in relation to book was increased slightly over the period but no other clear trend emerged. The incidence of Web citations began during the period but did not steadily increase. The study highlighted specific use patterns and collection weaknesses. (Leiding, 2005).

Hirsch (2005) developed the h-index which is an index that attempts to measure both the scientific productivity and the apparent scientific impact of a scientist. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other people's publications. Bornmann and Daniel (2005) found that on average the h-index for successful applicants for post-doctoral research fellowships was consistently higher than for non-successful applicants.

The h-b-index developed by Michael Banks (2005) is an extension of the h-index. Single h-index could not reflect the difference of time spans for scientists (Liang, 2006). The g-index was introduced as an improvement of the h-index (Egghe,
A significant correlation was found between the journals' h-indices and their citation impact scores (Saad, 2006).

To find the role of internet-based resources in citation works, a study was carried out by Megnigbeto (2006). The results revealed that Internet-based resources are very rare in the dissertations (less than 10%). A few students give references to the maximum of Internet-based resources. Cronin and Meho (2006) showed how the h-index can be used to express the broad impact of a scholar's research output and it is linear in career length (Burrell, 2007).

Huang (2007) proposed an integrated conceptual model to assess the journal citation impact. The result showed that the antecedent factors account for 39.4% of the variances in journal citations. The result of measurement model showed that the number of article is a good indicator for journal attributes; document availability is a good indicator for journal accessibility; and subject variety is a good indicator for journal diversity. As for the journal citation, three indicators that include impact factor, immediacy index, and total citations, are all good indicators. Moreover, the result of structural model showed that journal citation is significantly influenced by journal diversity.

Citation analysis was used as a tool to take decisions about how best to allocate acquisition expenditures among increasingly expensive journals. The study measured the return-on-investment of serial expenditures through the use of citation analysis (Via & Schmidle, 2007). A citation analysis of the journal articles was conducted to assess the needs of library collections and the evaluation by Rethlefsen (2007). The results showed that 63% of all citations were journal articles and most
cited materials were between 2 and 5 years old. The top 10 cited journals accounted for 36% of all citations and 62% of journals were cited only once.

Lotkaian informetrics theory was used to calculate the dynamical g-index (Egghe, 2007b). Cumulative $n^{th}$ citation distribution (Egghe, 2007a) and Maxprod index $h_2$ (Kosmulski, 2007) were discovered using citation analysis. A rational, successive g-index was proposed, and applied to economics departments in Ireland (Tol, 2008). The ability of g-index and h-index to discriminate between different types of scientists was analyzed in the area of Natural Resources at the Spanish CSIC. However, g-index is more sensitive than h-index in the assessment of selective scientists (Costas & Bordons, 2008).

Egghe and Ravichandra Rao (2008) calculated using those indices by following the two-by-two Spearman rank correlation coefficient and proved that these rankings are significantly related. In the discrete setting some small deviations from the ideal may occur (Egghe & Rousseau, 2008). Practical limitations of h-index in the content and software of the databases were illustrated by Jacso (2008a). Egghe (2008a) investigated the influence of arbitrary information production process transformations on the h-index and on the g-index and calculated concrete transformation formulae for the h-index (Egghe, 2008b). Bar-Ilan (2008) compared the h-indices in Web of Science, Scopus and Google Scholar. The pros and cons of the three largest, cited-reference-enhanced, multidisciplinary databases were found in another study (Jacso, 2008b).

Time series of h-indices were calculated in some specific models by Liu and Rousseau (2008). H-index of conglomerates was defined by Rousseau, Guns and Liu (2008). A bibliometric study was carried out on the citations within the chemistry
field Ph.D. dissertations to ascertain what types of documents are the most frequently used in the research process, the most frequently consulted journals and obsolescence rate of the journals. The study revealed that the most frequently used documents (79%) were scientific papers, journals met 50% of the informational needs; and the age of 50% of the citations was no older than 9 years (Vallmitjana & Sabate, 2008).

Jiao, Onwuegbuzie and Waytowich (2008) found out that levels of library anxiety predict simultaneously the citation error rate and quality of reference lists in doctoral dissertation proposals among 93 doctoral students in education. This unique study examined the bibliographic citation inaccuracies in doctoral research proposals and the current investigation was one of the first to investigate the psychological characteristics of doctoral students who commit such errors. A canonical correlation analysis revealed a multivariate relationship between levels of library anxiety and both the citation error rate and quality of reference lists.

A study by comparing among the journal citations data on education, library and information science, and management was carried out by Tsay (2008a). He used information in SSCI Journal Citation Reports to carry out this study. The correlation between each of the fifteen pairs of source items and five kinds of citation data-citation count, impact factor, immediacy index, citing half-life, and cited half-life-are examined based on the Pearson correlation tests. The significance of mean difference of each citation data was found by examined t test.

To find out the relationship of different types of citation data, a study was carried out by taking citation data from 53 journals of library and information science. It showed the relationship between each citation data and two self-citation rates, and the mean difference between two self-citation rates. All were examined by statistical
tests (Tsay & Chang, 2008). Using citation data from Journal Citation Reports (JCR) 1996-2004, a research was performed to show the relationship of citations usage between Library and Information and other fields (Odell & Gabbard, 2008).

To analyze and evaluate the subject change for both citing and cited literature on digital libraries based on bibliometric techniques a study was carried out by Tsay, (2008b). The results showed that most of the citing core journals on digital libraries are dedicated to the subject of ‘the application of computer and information technology to library implication’, while the cited literatures on digital libraries exhibit themselves mainly into four types of journals, namely, digital library orientation, general library and information science, new development in librarianship, and library technology.

An analytical study was carried out which focused on a citation analysis of research articles from scholarly electronic journals published in 2000-2006. The analysis also focused on the extent to which scholars are using web-based sources in scholarly electronic journals. Results of the study showed that 81.49% of articles have web references, 56.54 % of references are print journal references and 43.52% of them are web references (Bhat & Kumar, 2008).

A bibliometric study used 27 LIS dissertations with a total of 6,257 citations. The study result showed that most of the citations are from periodicals (42.2%), mostly from the US and the Indian journals. Nearly half of the citations (42.2%) are from the journal articles. Nearly three quarters of the journal articles were from the LIS journals, with the remainder 33 were from other subject areas. (Chikate & Patil, 2008).
The b-index was proposed by Brown (2009) for excluding self-citation from h-index values. A comprehensive review on the h-index and related indicators field was conducted by using different citation databases (Alonso, Cabrerizo, Herrera-Viedma & Herrera, 2009). The Eigenfactor score (Bergstrom, 2007), Article Influence score, and the Eigenfactor subject category list were briefly explained to discover the monetary value of research publications (Crisp, 2009). Wouters (1999) in his “the reflexive citation theory” explained that citations as indicators constituting a “formalized representation” of science that initially neglect meaning. According to this theory “this attribution of meaning can be postponed” and should be based not on the citing behavior of the citing scientists but on how citations reflect the characteristics of science.

To find the potential association between high citation counts and high quality research, Levitt and Thelwall (2009) investigated the 82 most highly cited Information Science and Library Science articles. The result showed that there is a moderate correlation (0.46) between citation ranking and the number of years between the peak year and the year of publication. It also cleared that the relative frequency of these 82 articles was much lower for articles solely in IS&LS than for those in IS&LS and at least one other subject, suggesting that the promotion of interdisciplinary research in IS&LS may be conducive to improving research quality.

To find the picture of print and electronic cited documents, an analytical study of citation analysis was conducted by Jan (2009). He showed that the highest number (52) of articles was published in 2004 in the Library Trends during the period from 1944-2007. The Journal contained 15662 references for the study period of which 13783 are p-citations and 1879 are e-citations. Every issue published approx 11 articles and each article had an average of 23.2 p-references and 3.1 e-citations. It was
found that 44.51% print books are consulted by the authors and 0% e-books are accessed. Mukherjee (2009) showed in his paper that articles with fewer citations mostly cite less-scholarly sources such as Web pages, whereas articles with a higher number of citations mostly cite scholarly sources such as journal articles.

Another study showed that citation analysis is a tool to select books for the social science book collection. It also indicated that books purchased using traditional methods of selection circulated more, except when individual disciplines were measured. Citation analysis methods accounted for circulation of nearly 95% of the social science collections (Enger, 2009).

Egghe (2009) indicated that if the Lorenz curve L(X) of X is above the Lorenz curve L(Y) of Y, then the g-index g(X) of X is larger than or equal to the g-index g(Y) of Y. This is a good property for so-called impact measures which is not shared by other impact measures such as the h-index. To show the frequencies of article impact in different abstracting databases Martell (2009) conducted a study on Yahoo, Google, Google Scholar, and ISI Web of Knowledge. Yahoo, Google, and ISI Web of Knowledge averaged between 2.8 and 3.5 citations per title for the period covered and Google Scholar averaged 6.4.

Egghe (2010) showed that Randic's H-sequence and H-index are equivalent in some situations. A research was undertaken to examine the correlation between h-index and rankings by peer assessment by Norris and Oppenheim (2010). Applications of h-index in ranking universities (Lazaridis, 2010) and French departments of Economics, departments of Management and Business schools (Courtauld, Hayek, Rimbaux & Zhu, 2010) were successfully applied.
To minimize the disadvantages of those indices, hg-index was developed taking all advantages h-index and g-index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2010). Schreiber (2010) analyzed the citation records of 26 physicists in order to determine the modified g index gm by fractionalized counting of the publications. The results of the correlations between these indices are relatively strong; the arrangement of the datasets is significantly different in detail depending on whether they are put into order according to the values of either the original or the modified indices.

An alternative approach based on the full structure of the scholarly citation network was initiated to describe Eigenfactor, Metrics-Eigenfactor Score and Article Influence Score-use metrics and the rankings that they provide (West, Bergstrom & Bergstrom, 2010). Some reasons were introduced to motivate the use of the Eigenfactor method (Franceschet, 2010). Using the h-index and Eigenfactor values a study provided a simple graphical representation of the journals to aid decision making processes with regards to scholarly research and scientific journal publications (Yin, Aris & Chen, 2010). Jacso (2010) showed that both the Eigenfactor Score (EFS) and the Article Influence Score (AIS) use a five-year target window in the algorithm to quantify the scholarly impact at the overall journal level and at the article level respectively.

A comparative analysis was performed for journal distributions, major subject themes, and authorship characteristics of these articles by Blessinger and Hrycay (2010). Thirty-two highly cited articles that were influential to scholarly communication in library and information science (LIS) in the latter part of the twentieth century were identified and examined in this study. To analyze citation behaviours of students in the LIS field a study was conducted during 2003-2008 in
Tehran. The study shows that students' citation behaviours are in favour of books, and Farsi e-sources are lesser used by LIS students than English e-sources. The total number of book citations is far more than the total number of other formats (Riahinia, 2010).

Bornmann and Daniel (2010) presented the citation speed index as a meaningful complement to the h index. Relationships between the journals download immediacy index (DII) and some citation indicators were studied. Results suggested that the DII can be regarded as an independent indicator, but that it also has predictive value for other indicators, such as a journal's h-index. (Wan, Hua & Rousseau, 2010).

To launch embedded library services, citation analysis was used as a tool to assess the effectiveness of the library instruction programme. The results showed that this methodology is useful for both the individual libraries hoping to assess their embedded librarian programs, and the researchers interested in exploring the overall effectiveness of this new mode of information literacy instruction (Clark & Chinburg, 2010).
2.1 Conclusion

The citations of the article are just the mirror images of the reference. An important stage of citation analysis was started when Garfield (1998) first introduced Science Citation Index in 1963. However, as Small (1998) has correctly observed, the SCI did not invent the citation, as Wouters (1999) seems to think, any more than the dictionary invented words. Thus, to understand the nature of the citation, it is required to understand the nature of the reference. And to understand the nature of the reference, it is needed to understand a theory of citing that explains why authors cite the way they do. Ergo, the full potential of citation analysis needs to be realized.

The above review ensures that there is no study regarding recency of citations of articles published in the field of Library & Information Science. This research study attempts to demonstrate the citation patterns and the age of citations by measuring recency of citations as well as articles published in the field of Library & Information Science.
References


Jacso, P. (2010). Differences in the rank position of journals by Eigenfactor metrics and the five-year impact factor in the ITJournal Citation Reports. *Online Information Review*, 34 (3), 496-508.


