An international multidisciplinary analysis of scholarly communication through investigating citation levels

An international multidisciplinary analysis of scholarly communication through investigating citation levels

Jonathan M. Levitt, B.A., M.Phil., M.Sc., M.A.

A thesis submitted in partial fulfilment of the
requirements of the University of Wolverhampton
for the degree of Doctor of Philosophy

October 2008

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or for any other purpose (unless previously indicated). Save for any express of acknowledgements, references and/or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

Signature

Date

Abstract

This thesis seeks to demonstrate that the new facilities of Web of Science (WoS) online can be used in new ways to enhance understanding of scholarly communication. It investigates four aspects of scholarly communication: characteristics of highly cited articles, citation levels of collaborative articles, citation levels of multi-disciplinary articles, and patterns of annual citation of highly cited articles. For the first two topics it investigates the WoS category of 'Information Science & Library Science' (IS&LS), whereas for the other topics it compares diverse WoS categories in science and social science. Although its main data source is WoS, its investigation of disciplinarity also uses Scopus.

The thesis finds: (a) Highly cited IS&LS articles tend to be multidisciplinary and cited late, but are not necessarily first-authored by influential IS&LS researchers, (b) Amongst un-cite IS&LS articles the proportion of collaborative articles has remained almost constant over the past three decades whereas for higher cited articles it has grown steadily with time, (C) In social science subjects the level of citation of multi-disciplinary research are generally similar to that of mono-disciplinary research, whereas in science the citations levels for multi-disciplinary research are substantially lower than that of mono-disciplinary research, and (d) In both science and social science many very highly cited articles continue to be heavily cited more than twenty years after publication.

This thesis also introduces and uses an indicator for measuring the extent of collaboration called 'average partner scores' and indicates a way in which the subject categories of WoS can be investigated without requiring a licence for the WoS database. Finally, it identifies and addresses some of the technical problems of using WoS online to investigate scholarly communication.

Table of Contents

List of Tables

List of Figures

Glossary

This glossary presents, in alphabetical order, some of the terminology that is used frequently in this thesis.

average partner score: The average partner score of a set of articles is defined as the average of the partner scores (defined below) of all articles in the set.

collaborative rate: The collaborative rate of a set of articles is defined as the proportion of articles that are collaborative (Gómez, Fernandez and Sebastian, 1999).

h-index: An abbreviation of 'Hirsch Index' (defined below).

Hirsch Index: The Hirsch Index h of a collection of documents is defined as the largest number of documents that are cited h or more times (Hirsch, 2005).

IS&LS WoS: The set of all documents in Web of science in the subject category of 'Information Science & Library Science'.

IS&LS: The sat of all documents in the Social Science Citation Index in the subject category of 'Information Science & Library Science'.

Normalised Hirsch Index: The Normalised Hirsch Index of a collection of articles is defined as the largest number of articles that are cited h or more times

$$h_{norm} = 100\, h^2 / n$$

where h and n are the h-index (defined above) and number of documents of the set.

partner score: The partner score of an article is defined as 0 if the article is by one author, 1 if by two authors, 2 if by three authors, and 3 if by more than three authors.

SCI: Thomson Scientific's Science Citation Index.

SSCI: Thomson Scientific's Social Science Citation Index.

WoS: WoS is an abbreviation of the Web of science, an online database produced by Thomson Scientific that consists of the Science Citation Index, the Social Science Citation Index, and the Arts & Humanities Citation Index.

Acknowledgements

I would like to thank my supervisor, Prof. Mike Thelwall for all his very helpful feedback, guidance and advice on my research. Throughout my studies at Wolverhampton University I have been struck by the speed at which Mike has provided rapid feedback that has been very incisive, perceptive and relevant. Mike has been an inspirational guide on my research, in identifying and helping me avoid potential problems in my research. As a source of advice Mike is both knowledgeable and unstinting in his willingness to help others. I have enjoyed very much having had Mike as my supervisor.

I would like to thank my mother Gertrude Levitt for having suggested that I study under Mike, for her proofreading feedback on my research and for her encouragement throughout my studies. She has been very willing to proofread drafts of my research and has never wavered in her encouragement. I would like to thank my brother Prof. Michael Levitt for his numerous conversations on my research and his help with software usage. I would like to thank my sister Dr. Ruth Rettie for some interesting conversations on my research.

I would like to thank Dr. Dave Wilkinson for having read my draft thesis and pointed out typing mistakes and all members of University of Wolverhampton Statistical Cybermetrics Research Group for their feedback on my research. I want to thank Prof. Dave Nicholas, Dr. Ian Rowlands, and Prof. Grant Lewison for their feedback on my early research. Finally, I want to thank my examiners Prof. Isabel Gómez and Dr. Peter Musgrove for their very helpful feedback.

Dedication

I would like to dedicate this thesis to all the people who have directly or indirectly helped me adjust to being visually challenged and, in particular, to my mother Gertrude.

Chapter 1: Introduction

1.1: Background

Scholarly communication is defined as "how scholars in any field use and disseminate information" (Borgman, 1989). Following on from this definition, investigating scholarly communication can enhance understanding of the use and dissemination of research. Research into scholarly communication can increase understanding of communication and inform decisions on the allocation of research funding.

In his seminal book 'Big science, Little science', Price (1963, p. 284) provided the following definition of a citation: "If Paper R contains a bibliographic footnote using and describing Paper C, then R contains a reference to C, and C has a citation from R." Citation analysis is the quantitative investigation of citations. One of the underlying assumptions of citation analysis is that high citation is associated with research quality. This association is not clear-cut; for instance, Moed (2005, p. 37) wrote, "Authors cite from a variety of motives, some of which may have little to do with research 'quality'." Nonetheless, as high quality research is often highly cited, citations are used to indicate research quality. Several potential uses of citation analysis were identified in the 1970s. Merton wrote that citation analysis can "trace the genealogy of scientific knowledge" (Garfield, 1979, p. viii) and Garfield (1979) suggested that citation analysis could be used for historical research into Science, to map the structure of science, and for journal evaluation.

The main data source of this thesis is Web of Science (WoS) online, a searchable database of over 36 million articles and other journal documents designated into over 200 subject categories. In common with Internet search engines, entering a WoS search query retrieves records and displays the reported number of matches; this thesis uses the reported number of matches to search queries as its data source. The search facilities of WoS became much more powerful in 2006, with the introduction of the 'Refine your results', 'Analyse', 'Create Citation Report' and 'Times cited' facilities; the 'Refine your results' facility was used in the data collection until it was withdrawn from WoS in February 2008. The 'Analyse' facility allows the user to obtain more detailed information on the search results; for instance the user can obtain details of the number of results in each subject category. Create Citation Report allows the user to identify the average number of citations in the search results, and the 'Times cited' facility sorts the search results into citation order, beginning with the most highly cited. The mechanism of these facilities is described in Section 4.1. WoS has been used to investigate diverse aspects of scholarly communication. Of particular relevance to this thesis is the past research on collaboration, disciplinarity and patterns of citation (described in Sections 2.2 to 2.4), as these are the aspects of scholarly communication investigated in this thesis.

Some influential information scientists have advocated combining bibliometrics with other methods. For example: (a) Borgman wrote: "In combination with data gleaned from other methods, they (i.e. bibliometric methods) can provide a large, rich characterization of communication processes not otherwise possible" (Borgman ed., 1990, p. 14), and Van Raan wrote "For a substantial improvement in decision-making, our bibliometric method has to be used in parallel with a peer-based evaluation procedure" (Cronin and Atkins ed., 2000, p. 303). Whilst it is appealing to use citation analysis alongside other methods, this is beyond the scope of this Ph.D.

1.2: Aim and objectives

This thesis aims to evaluate the extent to which the new facilities of WoS can be used to enhance the understanding of scholarly communication.

Its primary objectives are to:

- Use the new facilities to obtain findings in citation analysis of interest to the field of scholarly communication.
- Identify some of the opportunities to use the new facilities to investigate scholarly communication and to identify some of the more interesting potential applications.
- Identify some of the limitations of using WoS to investigate scholarly communication.

A secondary objective is to identify and address some of the methodological issues associated with using WoS to study scholarly communication.

1.3: Scope

This thesis investigates four aspects of scholarly communication: Characteristics of highly cited LIS articles (Chapter 5), citation levels of collaborative LIS articles (Chapter 6), citation levels of Multi-disciplinary articles (Chapter 7), and patterns of annual citation of highly cited articles (Chapter 8).

Two factors guided the choice of these investigations: Collectively they cover diverse aspects of scholarly communication in which there is recent research (described in Chapter 2) and individually they make strong use of the powerful new facilities of WoS which became available in 2006. These facilities enable each study to identify all the articles in a specific subject or combination of subjects and to rank them according to the number of citations. The facilities used in this investigation are outlined in the Methods.

In the studies of disciplinarity and late citation this research compares diverse subjects in both science and social science. The rationale for investigating both science and social science is that the findings would have greater generality. The study of characteristics of highly cited articles and the study of collaboration investigate Library and Information Science. The rationale is that for these studies it would have taken a disproportionate amount of time to investigate more than a single subject, and the subject category of 'Information Science & Library Science' (IS&LS) was chosen, as it is of particular interest to information scientists and there has been prior research on this subject category.

This research restricts itself to articles only and does not compare countries. Although the main data source is WoS, the investigation of disciplinarity also examines Scopus. One reason is to enable comparison of findings between Scopus and WoS; another reason (explained in the Methods) is that Scopus is more suited than WoS to the investigation of disciplinarity in science. Although the search facilities of EBSCO's Academic Premier became more powerful in 2006, this database was not used in this research, as it does not offer the facility to identify articles in subjects or subject combinations.

1.4: Contribution

This thesis produces several findings about scholarly communication, including: (a) Many of the first authors of very highly cited IS&LS articles have published few IS&LS articles (Chapter 5), (b) The level of collaboration in un-cited IS&LS articles did not increase between 1976 and 2004 (Chapter 6), (c) For many subjects, on average the mono-disciplinary articles are cited more highly than the multi-disciplinary articles (Chapter 7), and (d) Very late citation is not uncommon amongst very highly cited articles (Chapter 8).

This thesis also introduces and uses new methods and indicators that could be useful in other studies. For example: (a) It uses the WoS online subject categories (Chapters 5 to 8), thereby indicating that these subject categories can be investigated without requiring a WoS database licence, (b) It introduces and uses 'average partner scores' to compare collaboration (Chapter 6), and (c) It uses Normalised Hirsch Indexes to compare levels of citation (Chapter 7).

Finally, it identifies and addresses some of the technical problems of using the new WoS facilities to investigate scholarly communication.

1.5: Structure of the thesis

This thesis examines four global research questions in four studies presented in Chapters 5 to 8. Chapters 5 and 6 (derived from Levitt & Thelwall, in press 2008a and under review) investigate Library and Information Science only, whereas the other studies (derived from Levitt & Thelwall, in press 2008b and in press 2008c) investigate diverse subjects in science and social science. Each of these chapters includes a preliminary section discussing the specific methods applied in the study, then the findings, followed by a discussion, and finally conclusions. Chapter 2 first presents the research related to the thesis as a whole and then the research related to the individual studies. Chapter 3 describes the global research questions that shape the investigations of this thesis. It also describes the research design that presents the options encountered when implementing the research questions and the methods used by other researchers and why they were not used in this thesis. Chapter 4 examines methodical issues generic to two or more of the studies. Chapter 9 describes the main limitations. Chapter 10 re-examines the aim and objectives and suggests areas for further research.

Chapter 2: Literature review

This chapter reviews the literature that relates strongly to the methods and informetric investigations in this thesis, focusing on topics related to more than one investigation. Section 2.1 reviews research that uses the WoS subject categories, because the WoS subject categories contain the main data used in this thesis. Section 2.2 reviews research on collaboration (investigated in Chapter 6), Section 2.3, research on disciplinarity (investigated in Chapters 5 and 7), and Section 2.4, research on patterns of citation (investigated in Chapters 5 and 8). Section 2.5 reviews the literature on the indicators used in this thesis, on criticisms of citation analysis and the WoS database, and on alternatives to the WoS database.

2.1: WoS Subject categories

WoS subject categories have been delineated using the Journal Citation Reports (JCRs) and the categories of the WoS database. Both types of delineation are at the journal level in that all articles in a journal are in the same subject category (or categories); an investigation in Chapter 3 found very considerable overlap between the journals in the two forms of delineation. This chapter reviews research that delineates using WoS categories, as this is the method this thesis uses to delineate subjects (as justified in Chapter 3). The section reviews eight studies that are closely related to this thesis. It demonstrates the wide range of uses of WoS subject categories and is presented chronologically to indicate how these uses have evolved; an overall analysis then follows.

The next four paragraphs present eight studies that have used the WoS categories to examine the core topics of this thesis, namely collaboration, disciplinarity and patterns of annual citation. Qin, Lancaster and Allen (1997) examined eight subjects (Agriculture, Biology, Chemistry, Earth sciences, Engineering, Mathematics, Medical sciences, Physics) and found considerable variation in the level of collaboration and inter-disciplinarity (measured by the number of WoS categories of the citing articles). Morillo, Bordons and Gómez (2001) compared the degree of inter-disciplinarity of two WoS categories (Applied Chemistry and Polymer Science). Their indicators of inter-disciplinarity included the extent to which journals were assigned into multiple WoS categories and the percentage of citations and references outside the category.

Rinia, Van Leeuwen, Bruins, Van Vuren and Van Raan (2002) compared for 15 subjects (e.g., Biology, Chemistry, Mathematics, Physics) the number of references per article and the extent to which subject areas referenced one another. Rinia, Van Leeuwen and Van Raan (2002) examined the degree of inter-disciplinarity of research programs in Physics in the Netherlands. They defined an inter-disciplinarity Physics publication as a publication both in Physics and a WoS category not classified as Physics.

Morillo, Bordons and Gómez (2003) compared the extent of inter-disciplinarity of nine subjects (Agriculture, biology, and environmental sciences, Chemistry, Clinical medicine, Engineering/technology, Biomedicine, Physics, Humanities, Mathematics, Social sciences). Their measure of inter-disciplinarity was the extent to which the journals in the subject area were assigned to multiple WoS categories. Aksnes (2003) compared patterns of annual citation, a topic investigated in Chapters 5 and 8. One difference between Aksnes' study and this thesis is that Aksnes grouped together the WoS categories into subjects (Biology and environmental sciences, agriculture, Clinical medicine, Engineering, computing and technology, Life sciences, Physical, chemical and earth sciences) whereas this thesis examines individual WoS categories. Another difference is that Aksnes investigated articles by authors solely from Norway, whereas this thesis is not restricted to authors from only one country.

Zitt, Ramanana-Rahary and Bassecoulard (2005) compared the citation rates of different levels of grouping of journals, including the WoS categories and subject areas consisting of WoS categories. Adams (2005) investigated the extent to which early citation ranking correlates with late citation ranking, a topic examined in Chapter 8. In common with Aksnes, Adams grouped together the WoS categories into subject areas (Biochemistry and biophysics, Molecular biology and genetics, Optics and acoustics, Pharmacology and toxicology, Physical Chemistry, Space Science) and his study was confined to the U.K.

The above studies collectively have used the WoS categories to examine the core topics of this thesis. The diversity of these studies indicates both that the WoS can be used in a wide range of contexts and that there are considerable gaps in research on the WoS categories. This thesis focuses on one of these gaps, factors affecting citation, a subject addressed by Aksnes; however, a major difference between the research of Aksnes and this thesis is that this thesis does not confine itself to one country.

Using WoS subject categories is clearly an oversimplification, not least because Bradford's (1934) law of scattering implies that research is not always published in the core journals of a field, but also because some multi-disciplinary journals, such as Nature and Science, publish a significant amount of mono-disciplinary research (Ackerson & Chapman, 2003). Moreover, the subject categories of databases are optimised for information retrieval rather than for scientometric evaluation, and the issue of identifying disciplines is complex and without easy solutions (Glänzel & Schubert, 2003). Nevertheless, the simplification used here seems a reasonable method to differentiate between two sets of journals, one of which is likely to contain higher levels of inter-disciplinary research than the other. In addition, the assumption is supported by empirical data; for instance, Morillo, Bordons and Gómez (2001) found that a WoS subject with a high level of overlap with other subjects received proportionately more external citations than a WoS subject with a low level of overlap with other subjects.

2.2: Collaboration

In this thesis the word 'collaboration' is used to denote an article that has more than one author. Bordons, Gómez, Fernandez, Zulueta, and Mendez (1996) refer to three types of collaboration: local, domestic and international. In local collaboration all collaborators work in the same institution, in domestic collaboration not all collaborators work in the same institution but all collaborators work in the same country, and in international collaboration not all collaborators work in the same country.

Bordons and Gómez (2000, p. 197) describe some uses of bibliometrics in policy-relevant research on collaboration, stating that, "Hot topics on the agenda of research policy administrators, such as the collaboration between institutional sectors (i.e., triple helix issues concerning industry-university-government collaboration) or the establishment of cross-disciplinary collaborative links, can be undertaken using bibliometrics."

Section 2.2 focuses on research into links between collaboration and citation. Investigating collaborative articles does not fully capture the dynamic process of working together (Katz & Martin, 1997; Cronin, Shaw, & La Barre, 2003; Wang, Wu, Pan, Ma, & Rousseau, 2005), in that working together does not automatically result in collaborative publication. In addition the link between citation and 'quality', however conceived, is not clear-cut. None-the-less several studies have found it useful to investigate the relationship between collaboration and citation with a view to exploring the relationship between collaboration and quality.

Several macro-level studies have used the Science Citation Index (SCI) to investigate the relationship between collaboration and citation in science. Some studies have investigated the entire SCI whereas others have compared subject areas. In an investigation of the entire SCI for 1980-2000, Persson, Glänzel and Danell (2004) found a linear relationship between mean citation rates and the number of authors.

An investigation of nearly half a million U.K. publications from 1981 to 1994 (Katz & Hicks, 1997), found that articles by authors from two countries on average received about 50% more citations than articles by authors from a single country. Gómez, Fernandez and Sebastian (1999) found subjects varied considerably in their levels of international collaboration and Glänzel (2002) found a strong recent trend towards multiple authorship in biomedical research and chemistry. A link was found between higher citation and international collaboration for Brazilian Management Science (Pereira, Fischer, & Escuder, 2000). But in social science some studies that use the Social Science Citation Index (SSCI) have *not* found statistically significant correlations between collaboration and citation. These include investigations of sociology (Crase & Rosato, 1992), finance (Avkiran, 1997), ecology (Leimu, 2005), and two library science journals (Hart, 2007).

Several studies of international collaboration have also found positive associations between international collaboration and citation rates. These include investigations of Chilean physics (Vogel, 1997), Scandinavian science (Glänzel, 2000), Brazilian science (Leta & Chaimovich, 2002), New Zealand science (Goldfinch, Dale, & DeRouen, 2003), Danish industry (Frederiksen, 2004), HIV/AIDS in Nigeria (Uthman, 2008) and wood preservative chemical research (Yi, Ao, & Ho, 2008). In molecular biology, Ma and Guan (2006) found a correlation between collaboration and citation for Chinese molecular biology, whereas Herbertz (1995) did not find this correlation amongst well-known research institutes.

One concern, when investigating collaboration, is that articles by multiple authors seem more likely to be self-cited as there are more authors to self-cite. A study of international collaboration (Van Raan, 1998), however, concluded that higher rates of self-citation in international collaboration do not play a significant role in increasing the citation impact of internationally collaborative articles.

2.3: Disciplinarity

The beliefs that inter-disciplinary collaboration is conducive to quality in research and that some problems are too complex to be solved in a single discipline underlie the recent policy goal of encouraging collaboration between researchers in different disciplines, especially as part of modern applied inter-disciplinary 'Mode 2' research (Gibbons, Limoges, Nowotny, Schwartzman, Scott, & Trow, 1984). One perceived advantage of Mode 2 research is that it opens knowledge production to a wide range of influences (Leydesdorff & Etzkowitz, 2001).

Inter-disciplinarity has been encouraged in science policy both by creating multi-disciplinary centres and units and by funding multi-disciplinary research projects (Bordons, Zulueta, Romero, & Barrigon, 1999). Many science policy documents express high expectations of the benefits of inter-disciplinary research (Rinia, Van Leeuwen, Bruins, Van Vuren, & Van Raan, 2002). There has recently been a sharp rise in the number of policies and the amount of funding aimed at promoting cross-disciplinary collaboration between different fields, leading to claims that cross-disciplinarity has become the 'mantra of science policy' since the mid 1990s (Rafols & Meyer, 2007). In addition, an entire chapter of Moed (2005) discusses an example of a national Research Council seeking to stimulate trans-disciplinary research.

Recent years have also seen an increase in the use of citations for research evaluation, recommended for the U.K. after the 2007 Research Assessment Exercise (RAE) (http://www.hefce.ac.uk/research/assessment/faq/, last visited September 5, 2008). Previous U.K. RAEs have recognised concerns from multi-disciplinary researchers about the fairness of discipline-based peer evaluations of their work (RAE, 2004, paragraph 12) and hence it is increasingly important to understand the relationship between multi-disciplinarity and citation levels so that multi-disciplinary researchers are not unfairly disadvantaged – or advantaged – by citation-based metrics.

Inter-disciplinarity is now considered to be essential for the advancement of science (Bordons, Zulueta, Romero, & Barrigon, 1999) and several articles have analysed the perceived link between inter-disciplinarity and research quality. Suggested benefits of collaboration across discipline boundaries include: (a) Bringing multiple perspectives to bear on a problem, (b) Merging knowledge across disciplinary boundaries, and (c) Creating ways to address problems that cut across traditional fields of research (Haythornthwaite, 2006). Furthermore, inter-disciplinarity is considered the most effective way of addressing practical research topics (Morillo, Bordons, & Gómez, 2003).

Policies to stimulate inter-disciplinary co-operation are partly based on evidence from the history of science that breakthroughs in one field proved to be important for progress in other fields; for example, research on atomic spin in physics led to the development of magnetic resonance imaging that is used in medicine (Rinia, Van Leeuwen, Bruins, Van Vuren, & Van Raan, 2001). Calculus is another example of the transfer of knowledge; although developed for mechanics it is used in many other disciplines. Although knowledge transfer is used as an argument for inter-disciplinary co-operation, it is not clear-cut that inter-disciplinary research is conducive to knowledge growth. Specifically, progress in physics led to applications in medicinal diagnostics, but it is not clear-cut that this progress would have been fostered by inter-disciplinary co-operation between researchers in physics and in medicine.

Amongst previous quantitative investigations of disciplinarity, more than twenty-five years ago Le Pair (1980) examined the relationship between field mobility and the mutual influencing of different disciplines. One of the earliest citation analyses of inter-disciplinarity, that of Porter and Chubin (1985), found that citations across broad disciplinary categories were rare, although this has probably since changed. More recent investigations have obtained a number of qualitative findings. These include: (a) A few journals are mainly responsible for the cross-disciplinary citing of information science by the subject of ‘communication’ (Borgman & Rice, 1992), (b) The level of inter-disciplinarity varies considerably between disciplines (Qin, Lancaster, & Allen, 1997), and (c) Articles drawing information from a more diverse set of journals are cited particularly highly (Steele & Stier, 2000). A large-scale research into inter-disciplinary citations has shown that these tend to occur later than citations within the same discipline, with the exception of Mathematics and to a lesser extent Clinical Life Sciences (Rinia, Van Leeuwen, Bruins, Van Vuren, & Van Raan, 2001).

Citation analysis has been a useful tool for investigating disciplinarity, particularly in the context of examining the patterns of inter-disciplinarity in various fields (e.g., Herring, 1999; Pierce, 1999; Van Leeuwen, & Tijssen, 2000; Rinia, Van Leeuwen, Bruins, Van Vuren, & Van Raan, 2002; Rinia, Van Leeuwen, & Van Raan, 2002; Leydesdorff, 2007; Rafols & Meyer, 2007). Inter-disciplinarity is widespread in the sense that 25% of WoS journals are classified in more than one discipline (Rinia, Van Leeuwen, Bruins, Van Vuren, & Van Raan, 2002); for Scopus 27.3% of all articles in Medicine journals (the category with most articles) published before 2007 are classified in more than one subject. Nevertheless, few previous studies have investigated how citation levels vary with disciplinarity and there is no clear overall pattern. Another study included relevant data but did not directly address the issue for Information Systems research (Chan, Kim, & Tan, 2006). A previous study addressed the issue of whether multi-disciplinary research was more highly cited than mono-disciplinary research, using the case of physics in Dutch universities, and found slightly fewer citations per paper for multi-disciplinary research (Rinia, Van Leeuwen, Bruins, Van Vuren, & Van Raan, 2002).

Previous research has addressed the issue of inter-disciplinarity in IS&LS research. Rice and Crawford (1992) identified some possible areas of convergence between the fields of communication and library and information science. Meyer and Spencer (1996) found that articles in library science were cited in computer science, medicine, psychology, the social sciences, and general sciences whereas Tang (2004a) found that Information and Library Science "attracts a significant wide spectrum of disciplines from the domains of science, social science, and the humanities, and that the kinds of disciplines interested in the field vary by year."

Other research on disciplinarity and Information and Library Science include that of Carlin (2003) and Tang (2004b). Carlin conducted text analysis on papers by Elfreda Chatman in library and information science journals, and Tang found that there was considerable cross-disciplinary citation to and from 150 randomly selected LIS articles.

2.4: Patterns of citation

Several studies have quantified the late citation of highly cited articles. Aversa's (1985) investigation of the patterns of citation of 400 papers published in 1972 that were cited 30 or more times between 1972 and 1977 found that an early rise in being cited is associated with a more rapid decline in citation and a lower citation total, whereas a delayed rise in citation is associated with a less rapid decline in citation and a higher citation total. McCain and Turner (1989) compared four slowly ageing and seven rapidly ageing highly cited articles in Molecular Genetics. Regarding the findings, McCain (2007) wrote, "Late-peaking, slowly aging papers were likely to be cited for important widely useful methodologies or fruitful, broadly relevant theoretical insights while the early-peaking, quickly-aging papers were being cited primarily for their explicit research results." Similarly, Cano and Lind (1991) compared the annual citation patterns of ten highly cited papers with ten papers with low citation in medicine and biochemistry. They found two types of citation patterns, A and B. For articles of Type A the number of citations in the first six years, as a percentage of total citations, was typically 75% whereas for Type B the figure was typically 33%. Six of the highly cited, but none of the low cited, were of Type B. All six were classified as "sharing a methodological nature." Aksnes' (2003) investigation of the patterns of annual citation of 137 highly cited papers in Norwegian science published between 1981 and 1989 found that 33% of the papers in Physical, Chemical and Earth Sciences had the citation pattern of 'Early rise & rapid decline', whereas none of the papers in Biology and Environmental Sciences had this citation pattern. One pattern of citation referred to by both Merton (1968) and Garfield (1975) is the 'obliteration phenomenon'; Garfield (1975) used this phrase to describe basic findings that have become so widely recognised that they are used without citing their source. Garfield (1993) used the phrase 'obliteration by incorporation', in which "discoveries or ideas become so fully incorporated into canonical knowledge that their source is no longer cited or even alluded to." In contrast, Glänzel, Schlemmer and Thijs (2003) and Van Raan (2004) established the frequencies of some unusual citation patterns of highly cited articles which they termed respectively 'delayed recognition' and 'sleeping beauties'. Glänzel, Schlemmer and Thijs found that .3% of papers in the 1980 Science Citation Index that were cited more frequently than 15 times in total were not cited between 1980 and 1984; Van Raan found that only 41 of the articles from the Institute for Scientific Information Citation Indexes published in 1988 received at most ten citations during the first ten years after publication and subsequently between 21 and 30 citations in the next four years. Other studies of delayed recognition include those of Garfield (1980) and Glänzel and Garfield (2005) and other studies of sleeping beauties include those of Van Dalen and Henkens (2005) and Burrell (2005).

Other investigations of citation patterns have also produced interesting results. Garfield (1985a) presents a graph that compares patterns of three highly cited papers and Garfield (1985b) presents a graph that compares patterns of four highly cited papers. Garfield's graphs contain at least two different citation patterns: (a) Rising to a peak and then a steady decline, and (b) Continuing increase in citation level. Whilst previous research has investigated lateness of citation in different subject areas, none has examined late citation amongst the most highly cited articles in IS&LS. Chapter 5 examines late citation in IS&LS and Chapter 8 compares late citation in six subjects in science and social science.

2.5: Other research

2.5.1: H-index and related indicators

The h-index of a collection of documents is defined as the largest number (h) of documents that are cited h or more times (Hirsch, 2005). As described below, the h-index has been used as an indicator of research quality. One advantage that h-index has, compared with other indicators of research quality (e.g., the g-index, Egghe, 2006), is that the h-index is easy to obtain; the h-index of the results of a WoS search query can be obtained by using the sort by 'Times Cited' facility to rank the results in order of decreasing citation or from citation reports created using the 'Create Citation Report' facility.

The h-index has been used to compare influential information scientists. Cronin and Meho (2006) in an investigation of 31 influential information scientists from the U.S.A. found a Spearman correlation of .9 between their h-indexes and total number of citations. Cronin and Meho (2007) subsequently examined, for 12 renowned information scientists, links between age, the number of years as information scientists, and the publication of highly cited articles. Oppenheim (2007) used the h-index to rank influential British researchers in library and information science.

Studies have compared the h-indexes of journals (Braun, Glänzel, & Schubert, 2006) of countries (Csajbok, Berhidi, Vasas, & Schubert, 2007), and of researchers in different fields (Iglesias & Pecharroman, 2007). Van Raan (2006) compared h-indexes with other bibliometric indicators and peer review, and Minasny, Hartemink and McBratney (2007) examined the relationship between h-index and career length. Batista, Campiteli and Kinouchi (2006) used a modified h-index that takes into account the collaborative level of the field and Arencibia-Jorge, Barrios-Almaguer, Fernandez-Hernandez and Carvajal-Espino (2008) used a series of h-indexes at successively higher levels of aggregation, a method proposed in Schubert (2007).

Chapter 5 examines the h-indexes of first authors of highly cited IS&LS articles. The Normalised Hirsch Index is derived from the h-index and is used in Chapter 7 to compare citation levels.

2.5.2: Criticisms of citation analysis and the WoS database

An important criticism of citation analysis is that not enough is known about the citation process; to quote Cronin (1984, p. 86) "Citation is not something which happens in a void, and citations are not separable from the contexts and conditions of their generation … Future studies of citation should therefore concentrate of the content of citations, and the conditions of their creation and application." Moreover Moed (2005, p. 80) confirms this, "Authors cite from a variety of motives, some of which may have little to do with research 'quality'." Whilst it seems important to understand more about the significance of citation, large-scale studies of the significance are likely to be laborious.

A second criticism is that citation analysis makes assumptions regarding the nature of citation. For example: (a) Citation implies use, (b) Citation is based on merit, and (c) Citation reflects similarity of content, and (d) All citations are equal (Pierce in Borgman ed., 1990, p. 49). In response, although it is important when using methods to be aware of methodological assumptions, all methods have their strengths and weaknesses; as Moed (2005, p. 23) wrote "Not only citation analysis but also peer review of research performance has its particular strengths and limitations."

A third criticism concerns the coverage of the WoS database. For example: (a) They cover mainly North American, Western European, and English-language titles, (b) They are limited to citations from 8,700 journals, (c) They do not count citations from books and most conference proceedings, (d) They provide different coverage between research fields, and (e) They have citing errors, such as homonyms, synonyms, and inconsistency in the use of initials and in the spelling of non-English names (Meho & Yang 2007). The first three criticisms concern the scope of the WoS, and it can be important to take this into account when interpreting findings. The fourth criticism reflects that citation analysis is more suited to some fields (e.g., science) than others (e.g., humanities). The fifth criticism, of the inaccuracies in the data, indicates that it can be important not to assume that the data in citation databases is accurate.

2.5.3: Alternatives to the WoS database

Norris and Oppenheim (2007) found the facilities for record processing of the WoS and Scopus to be stronger than that of CSA Illumina and Google Scholar. This finding is consistent with Meho and Yang (2007), who found that the collecting the citations on the research of 25 LIS academics took 100 hours for

WoS, 200 hours for Scopus, and 3000 hours for Google Scholar. These findings correspond to the rationale in Section 3.2 for using the WoS and Scopus databases (in that they provide refined search facilities).

Meho and Yang (2007) found that using Scopus and Google Scholar, in addition to WoS, provided a more comprehensive picture of the citation impact; the investigation of disciplinarity (Chapter 7) complements the findings for WoS with those for Scopus. They found that Google Scholar's coverage of conference proceedings and non-English language journals is particularly strong, indicating that Google Scholar is particularly suited to those types of investigation.

Chapter 3: Global research questions and research 0design

This chapter consists of three sections: Section 3.1 presents the four research questions related to the aim and objectives (presented in Section 1.2), Section 3.2 describes the major decisions made regarding the implementation of these research questions and the rationale for deciding against alternative ways of proceeding, and Section 3.3 links Chapter 3 to subsequent chapters.

3.1: Global research questions

As stated in Section 1.3, this thesis focuses on collaboration, disciplinarity and patterns of annual citation, as they are of recent research interest; the global research questions focus on these topics. This thesis has four global research questions investigated in Chapters 5 to 8:

- **Question 1:** What are the characteristics of the most highly cited articles in Library and Information Science?
- **Question 2:** In general, are collaborative articles more highly cited than non-collaborative articles in Library and Information Science?
- **Question 3:** In general, are articles in multi-disciplinary journals more highly cited than articles in mono-disciplinary journals?
- **Question 4:** In general, do very highly cited articles have different patterns of annual citation from less highly cited articles?

The research questions investigate characteristics of highly cited articles, collaboration, disciplinarity and patterns of annual citation. These topics were chosen, as collectively they cover aspects central to scholarly communication in which there has been, and continues to be, extensive research interest. The central role of collaboration and disciplinarity in scholarly communication is indicated by the large amount of informetric literature on these topics and that the ISSI 2007 conference had sessions on both these topics. Although there has been less research interest in the characteristics of highly cited articles and their patterns of annual citation, these topics were chosen as their investigation can enhance understanding of the nature of research quality. In order to increase coherence the same subject was investigated in Questions 1 and 2, and both science and social science were investigated in Questions 3 and 4.

The rationale for examining whether collaborative articles are more highly cited than non-collaborative articles, is that many governments are promoting more international collaboration in the belief that this will result in reduced costs and higher impact research (Katz & Hicks, 1997). The rationale for examining whether multi-disciplinary articles are more highly cited than mono-disciplinary articles, is that inter-disciplinarity has been encouraged in science policy both by creating multi-disciplinary centres and units and by funding multi-disciplinary research projects (Bordons, Zulueta, Romero, & Barrigon, 1999). In addition, many science policy documents express high expectations of the benefits of inter-disciplinary research (Rinia, Van Leeuwen, Bruins, Van Vurena, & Van Raan, 2002). The rationale for examining whether patterns of annual citation of very highly cited articles differ from those of less highly cited articles, is that late citation amongst highly cited articles indicates that they have citation impact many years after publication. Several studies found late citation amongst highly cited articles (e.g., Aversa, 1985; McCain & Turner, 1989; Cano & Lind, 1991; Aksnes, 2003).

3.2: Research design

The research questions are addressed by a series of quantitative studies using commercial citation indexes, such as the Science Citation Index and Scopus, to generate statistics about article citations. In each study, subjects are delineated using the relevant subject categories of the citation index. This section discusses points of general relevance whereas for each study specific details of the research design are given in the relevant chapter.

The following decisions were made on the design of this research:

- **Decision 1:** To use bibliometric indicators (citation counts) rather than peer review to identify articles to be investigated.
- **Decision 2:** To use WoS as the primary data source rather than Scopus (or any other data source).
- **Decision 3:** To delineate subjects using the subject categories of WoS online rather than Journal Citation Reports (JCRs) or other methods (e.g., delineation on an article-by-article basis, as in Glänzel, Schubert, & Czerwon, 1999). Gómez, Bordons, Fernandez and Mendez (1996) compared some of the diverse ways of delineating subjects.

- **Decision 4:** To restrict science subjects to the SCI and social science subjects to the SSCI, as opposed to delineating subjects on the entire WoS.
- **Decision 5:** To investigate articles only.

Decision 1 is to use bibliometric indicators of research quality and disciplinarity. The rationale for not using peer review is that this research is a large-scale investigation and it would have been impractical, in a Ph.D. project, to arrange for experts in numerous subjects to assess research quality or disciplinary categories. It is reasonable to use bibliometric indicators because they provide findings at the subject and journal level to which tests for statistical significance can be applied.

Decision 2 is to use WoS as the primary data source rather than Scopus or Google Scholar. Both WoS and Scopus provide more refined search facilities than Google Scholar; Google Scholar investigations can be impractical due to the extent of manual filtering needed (Meho & Yang, 2007). WoS has the advantage over Scopus that WoS has been used much more extensively in research than Scopus and therefore there is more known about its imbalances. Another reason is that WoS provides citation data for a particularly long period of time; Scopus only has citation data for citations received after 1994 whereas WoS provides citation data on a large number of articles published as far back as the 1970s. WoS thus enables this research to tailor the citation window to the research question, rather than to confine the citation window to the period from 1994. For instance, the citation window for research question 4 is as much as 36 years as this very long period provides more information on patterns of annual citation.

There are other advantages of using WoS: (a) Its journal selection is more transparent and open to feedback in that ever since 1975 WoS has made available lists of journals in subjects through the Journal Citation Reports (JCRs), (b) Its coverage of social science is more comprehensive than Scopus (88% more articles in 1995), and (c) It is far more researched than either Scopus or Google Scholar and therefore new findings can be interpreted in the light of previous research. In addition, according to a personal communication by McVeigh in 2006, the categories are assigned by the ISI staff on the basis of a number of criteria, including the journal's title and its citation patterns (Leydesdorff, 2008). Some criticisms of the WoS database were presented in Section 2.5.2.

Decision 3 is to delineate subjects using the subject categories of WoS online. Other research that uses WoS subject categories includes Aksnes (2003), Adams (2005), Zitt, Ramanana-Rahary and Bassecoulard (2005) and Porter, Cohen, Roessner and Perreault (2007). Although other research has used subject categories either defined by experts or by other criteria, such as citation patterns, neither of these choices is practical for the large-scale investigations here. One reason for using WoS is that WoS was found to be more comprehensive than the JCRs. Specifically: (a) The IS&LS subject category contains over 16,000 articles published prior to the year of the earliest JCR (1975), and (b) For the period 2000 to 2006 the set of journals obtained by the first method contains all the journals obtained by the JCRs and 10 additional titles. These ten titles are presented in Table 3.1.

Table 3.1: Titles in IS&LS with articles published in 2000-2006 that are not in any of the JCRs for 2000-2006

Journal title
ASIST Monograph Series
Health Information And Libraries Journal
International Forum On Information And Documentation
International Information & Library Review
Journal Of Computer-Mediated Communication
Journal Of Education For Library And Information Science
Journal Of Global Information Management
Profesional De La Informacion
Publishing Research Quarterly
Web Of Knowledge - A Festschrift In Honor Of Eugene Garfield (edited book)

Omitting the articles published prior to 1975 would have resulted in three of the ten most highly cited IS&LS articles (including Henry Small's most cited article on co-citation analysis) being omitted from the investigation in Chapter 5; omitting the 10 extra journals of WoS for 2000 to 2006 would have had very little effect on the findings in this thesis, as only Chapter 6 examines IS&LS articles published during that period and the 10 journals contain only 3.9% of all IS&LS articles published in 2000 to 2006. Another reason for using WoS online is that the research was not restricted to the time periods for which the JCRs are available (JCRs are available online from 2000).

Some citation research uses the Dialog interface to Thomson Scientific/ISI data (e.g., White & McCain, 1998; Ingwersen, Larsen, & Wormell, 2000; Clausen & Wormell, 2001; White, 2001; Rousseau & Zuccala, 2004; Cronin & Meho, 2006; Stock & Stock, 2006), but this was decided against in order to keep within the thesis's goal of exploiting the standard WoS interface that is universally available in the U.K. (in contrast to Dialog). Dialog's facility to search references (Rousseau & Zuccala, 2004) may have provided the option of comparing the findings on references with the findings on citations; its facility of building large data sets that can be manipulated (Tenopir, 2001) may have allowed more in depth manipulation of data. Delineation conducted on an article-by-article basis (Glänzel, Schubert, & Czerwon, 1999) or by examining the terminology in article titles (Lewison & Paraje, 2004) are impractical in this large-scale investigation.

Decision 4 is to restrict, apart from the two exceptions described in the next paragraph, science subjects to the SCI and social science subjects to the SSCI. One reason for this decision is that previous research of subject categories has used JCRs to delineate subjects (e.g., Moed, Van Leeuwen, & Reedijk, 1999; Shama, Hellgardt, & Oppenheim, 2000; Pudovkin & Garfield, 2002; Leydesdorff, 2004; Leydesdorff, 2007) and JCRs confine science subjects to the SCI and social science subjects to the SSCI. Another reason is that, especially for social science, it would have required a disproportionate amount of time to have investigated the entire WoS database rather than solely the SSCI and, from data on the annual percent of IS&LS articles that are in SSCI (Table 3.2), 93.7% of all WoS articles in IS&LS published prior to 2007 are in the SSCI. In Table 3.2, 'SSCI' denotes the number of IS&LS articles in the SSCI, 'Not SSCI' the number of IS&LS articles not in the SSCI, 'All WoS S' the number of IS&LS articles in the entire WoS, and "SSCI as a % of all WoS' the percentage of IS&LS articles in entire WoS that are in the SSCI.

Table 3.2: Annual data on IS&LS articles in the SSCI and the entire WoS (November 2007)

Period	SSCI	Not SSCI	All WoS	SSCI as a % of all WoS
Before 1945	0	0	0	NA
1945-49	0	88	88	0
1950-54	0	76	76	0
1955-59	2,613	67	2,680	97.5
1960-64	3,663	270	3,933	93.1
1965-69	4,464	19	4,483	99.6
1970-74	4,969	310	5,279	94.1
1975-79	7,802	696	8,498	91.8
1980-84	8,417	1,176	9,593	87.7
1985-89	8,875	1,041	9,916	89.5
1990-94	9,845	1,352	11,197	87.9
1995-99	11,003	92	11,095	99.2
20000-04	11,075	1	11,076	100.0
2005-06	4,433	1	4,434	100.0
All	**77,159**	**5,189**	**82,348**	**93.7**

The two exceptions to Decision 4 both delineate IS&LS using the entire WoS and in order to indicate when IS&LS spans the entire WoS, the notation 'IS&LS WoS' is used. The two exceptions are: (a) In Sections 4.3 and 5.3 in the percentages of citing documents, and (b) In Sections 5.3.2 and 5.3.3 in the h-indexes and number of articles of influential information scientists. The rationale for exception (a) is the considerable technical difficulties entailed in limiting the citing documents to IS&LS; the rationale for exception (b) is that this study is seeking to demonstrate that a high percentage of influential information scientists have low h-indexes in IS&LS evaluated on the entire WoS, as opposed to solely on IS&LS restricted to the SSCI.

Decision 5 is to investigate articles only. The rationale is that there is more research interest in articles than in other types of documents.

3.3: Concluding remarks

This chapter describes the global research questions and issues of research design and Chapter 4 provides additional relevant details. Chapter 5 investigates research question 1, Chapter 6 research question 2, Chapter 7 research question 3, and Chapter 8 research question 4. Chapter 4 describes methodological issues relating to the research design and also the methods used in two or more of the investigations in Chapters 5 to 8.

Chapter 4: Methods

This chapter presents the methods used in more than one chapter; methods used in a single study are presented in the chapter containing the study. This chapter consists of six sections. Sections 4.1 to 4.4 focus on different aspects of the data collection: Section 4.1 describes the search features most central to the data collection, Section 4.2 describes ways in which the accuracy of the data was checked, Section 4.3 explains how subject categories were delineated, and Section 4.4 reports how the articles by notable influential scientists were identified. Finally, Section 4.5 describes the methods of data analysis and Section 4.6 links Chapter 4 to subsequent chapters.

The data used in this thesis is the reported matches of search queries in WoS and Scopus. This data was collected by manually entering searches rather than automatic queries, which are not permitted on WoS. The data was collected using the standard Web interfaces of WoS and Scopus, accessed through the electronic resources of Wolverhampton University and the Open University.

4.1: Database features used in the data collection

This section describes the WoS user interface that was introduced in February 2008. Although it is possible for the user to conduct searches directly from the login page, some of the key features used in this thesis (e.g., sort by 'Times Cited' and obtain a 'Citation History') are not offered from the login page. In order to access these facilities and to limit searches to the Science Citation Index or the Social Science Citation Index, the 'Select a database' button on the login page must be pressed, then the 'Web of Science' link, and finally the 'Change Limits and Settings' link. To illustrate the search process, in order to search for articles in the 'Scientometrics' journal, 'Scientometrics' is typed in the top text field box, 'Publication Name' selected from the top drop-down menu and the 'Search' button pressed. Figure 4.1 presents a screenshot of the displayed page.

Figure 4.1: Screen-shot of a search for 'Scientometrics' in the Publication Name (22nd February, 2008)

Figure 4.1 contains three features used extensively in the data collection for this thesis: (a) The ‘Analyse Results’ link, (b) The capacity to sort by ‘Times Cited’ (by altering the ‘Sort By’ drop-down menu from its default value of ‘Latest Date’), and (c) The ‘Create Citation Report’ link. These features, introduced by WoS online in Spring 2006, dramatically increase the range of data that can readily be retrieved. The next three paragraphs describe the functionality of these facilities and also of the ‘Refine your results’ facility that was available in WoS prior to the change in user interface and which was used extensively in the data collection. Although the current user interface has a ‘Refine Results’ section, it seems to be less powerful than the ‘Refine your results’ facility.

‘Refine your results’ and ‘Analyze Results’ both have eight options: Subject Categories, Publication Years, Countries/Territories, Institutions, Source Titles, Document Types, Authors and Languages. The studies in Chapters 5 to 8 repeatedly use the Subject Categories and Publication Years options. As a prelude to investigating the informetric properties of articles within a subject, the studies identify the journals that contain articles in the subject; in order to identify these journals the Subject Categories option is used extensively. The studies frequently examine the patterns of annual citation of articles; in order to identify these patterns the Publication Years option is used extensively. One advantage of ‘Analyze Results’ is that it lists up to 500 results, whereas ‘Refine your results’ only lists up to 100; a disadvantage is that on occasion the mechanism on ‘Analyze Results’ for restricting the search results did not function effectively.

'Sort by' offers the option of sorting search results according to Latest Date, Times Cited, Relevance, First Author, Source Title, and Publication Year. The second option is particularly useful, as it enables search results to be ranked according to the number of citations. Chapters 5 to 8 investigate the extent to which informetric findings depend on the level of citation; in order to isolate the papers at different levels of citation the 'Times Cited' option is used repeatedly.

'Create Citation Report' provides data on the citations. The 'View' option enables information on the citations to be obtained using the 'Refine your results' and 'Analyze Results' and data obtained using these options was extensively used in the investigations of patterns of annual citation (Chapters 5 and 8).

WoS limit the number of documents that can be processed, making the task of processing a large number of records (e.g., isolating all IS&LS articles) more arduous. The Refine your results, Analyze Results and Times Cited facilities can be applied only to the results of searches, and irrespective of the number of matches to the search query, a maximum of 100,000 are isolated in a search query. The impact of this limitation is that the Refine your results, Analyze Results, and Times Cited facilities can only be applied to a maximum of 100,000 documents. The Create Citation Report facility has a stronger limitation, in that currently it can only be applied to a maximum of 10,000 documents at one time. Another limitation is that searches of WoS require at least one text field to contain text. Section 4.3 describes the impact of these limitations on the task of isolating all articles in a subject category.

Many of the features of WoS are also available on Scopus. The untitled page reached after implementing a search, offers several features used in the investigation of disciplinarity (Chapter 7). The features used include the option of restricting search results to specific subjects ('Subject Area' of 'Refine Results', and of ranking the results of searches by citation level 'Cited By' button). These facilities do not limit the number of documents that can be processed. Scopus has the limitation that the Basic Search requires that at least one text field contains text; this limitation can be circumvented using the Advanced Search facility (e.g., by entering 'test' in the 'Search for' text box of the Basic Search and deleting TITLE-ABS-KEY(test) in the Advanced Search).

4.2: Reliability of the data

As the validity of the findings depends critically on the reliability of the data, before deciding to use WoS for research purposes extensive checks were conducted on the reliability of the reported number of matches.

The following checks were implemented on WoS before deciding to use the reported matches as a data source:

Check 1: Are the reported matches for an interval of several years identical to the sum of the reported matches for the individual years? For example, are the reported matches for 'USA' for the period 2005 to 2006 identical to the sum of the reported matches for 'USA' for the years 2005 and 2006?

Check 2: Do the reported matches depend on the order in which the search terms were entered? For example, are the reported matches for 'JASIST OR UK' identical to the reported matches for 'UK OR JASIST'?

Check 3: Are the reported matches identical to the actual number of matches? For example, are the reported matches for 'Thelwall Wolverhampton JASIST' equal to the actual number of matches for that search query?

The rationale for implementing Checks 1 to 3 is that significant inaccuracies in the reported matches of Google search engines were found using these checks. An example of an inaccuracy revealed by Check 1, is that on 2nd December 2007 for Google Scholar the reported matches of 'USA' for 2005 to 2006 were less than a quarter of the sum of the reported matches for 2005 and 2006 (201,000 for 2005-06, 489,000 for 2005 and 369,000 for 2006). An example of an inaccuracy revealed by Check 2 is that on 2nd December 2007 for Google Scholar the reported matches of 'JASIST OR UK' for 2006 was 22,000 more than that for 'UK OR JASIST' (316,000 compared with 294,000). An example of an inaccuracy revealed by Check 3 is that on 2nd December 2007 for Google the reported matches of 'Thelwall Wolverhampton JASIST' was more than double the actual number of matches ('about 134' compared with 56).

Every few months, checks 1 to 3 were conducted on WoS data and no inaccuracies were detected. In addition, the researcher implemented the following checks on the WoS data used in this research:

Check 4: Do the reported matches depend on the manner in which the task was specified? For example, WoS allows time periods to be specified in three different ways; are the reported matches identical for the different ways of specifying time periods?

Check 5: Is the content of searches accurate? For example, does a search for JASIST articles in 2006 solely locate articles, published in JASIST in 2006?

Check 6: Do the reported matches vary with time? For example, is the reported number of articles published in a fixed period in the subject of Information Science & Library Science (IS&LS) constant?

Check 4 identified one source of inaccuracy in WoS data, in that the reported matches can depend on the way the time period is specified. For example, the reported matches for articles in the Library Journal for the SSCI for 2005 had at least three values (116, 125 and 138) that depended on the manner in which the search was conducted; 116 was obtained when '2005' was selected in the 'Year' drop-down menu and '2005' typed in the 'PUBLICATION YEAR:' text-box, 125 was obtained when '2005' was selected in the 'Year' drop-down menu and nothing was typed in the 'PUBLICATION YEAR:' text-box, and 138 was obtained when '1900-1914' and '2008' were selected in the 'From' and 'To' drop-down menus and '2005' typed in the 'PUBLICATION YEAR:' text-box. Sorting the results by date revealed that all 116 articles from the first search were dated 2005, 116 articles from the second search were dated 2005 and the other 9 dated 2004, and all 138 articles from the third search dated 2005. In addition to illustrating Check 4, this data also illustrates that the content of searches on WoS can be inaccurate in that 9 of the matches for the second search were not published in 2004. Fortunately this problem does not seem to be widespread in that the only example of inaccuracy that the researcher identified was when the year of publication was not typed in the 'PUBLICATION YEAR:' text-box. In addition, where possible this research uses the last method of delineating time periods, as it identified the most results from 2005 and did not have any results from other years. The only exception is in Chapter 5, where for technical reasons, it uses the second method when isolating the IS&LS articles published prior to 2007.

An example of the reported matches of WoS varying with time (Check 6) was identified when checking the number of WoS articles in IS&LS published before 2007. The number of IS&LS articles evaluated in August 2007 was 373 higher than that for February 2007. A total of 109 of these additional articles were published in the Journal of Computer-Mediated Communication (62 in 2005 and 47 in 2006), 93 in Health Information and Libraries Journal (49 in 2005 and 44 in 2006), 92 in Rairo-Informatique-Computer Science (21 in 1977, 19 in 1978, 18 in 1979, 18 in 1980 and 16 in 1981), 51 in Profesional de la Informacion (all in 2006), and 28 in Journal of Global Information Management (15 in 2005 and 13 in 2006). These findings indicate that WoS added five IS&LS journals between the dates of the two studies, and that three of the journals had articles for 2005 and one for as far back as 1977. This indicates that WoS holdings are not fixed and can alter retroactively. The effect is relatively minor, though, as the 373 extra articles are less than half a percent of over 82,000 articles in IS&LS.

The checks on Google indicate that for Google Scholar and Google the reported matches is not a reliable data source. Although the checks on WoS revealed two different sources of inaccurate data, the following ways were found to cope with them: (a) Where possible specify periods by typing the period in the 'PUBLICATION YEAR:' text-box and using the default values in the 'From' and 'To' drop-down menus, and (b) Keep a record of when the data was collected, as the holdings of WoS can change retroactively. The testing of Scopus was less thorough than that for WoS, as Scopus data was used in only one study and WoS data was used in four studies. The researcher conducted Checks 1 to 3 on Scopus and did not identify any inaccuracies.

4.3: Delineation of subjects

The delineation of subjects is fundamental to Chapters 5 to 8, in that each chapter investigates one or more subject that needs to be delineated in this research. As mentioned in Section 3.2, this research delineates subjects using the subject categories of WoS online; the limitations described in Section 4.1 can make this task laborious.

Three limitations impinge on this task: (a) The 'Refine your results' and 'Analyze Results' facilities can be applied only to the results of searches (so subject categories cannot be specified on the search page), (b) Searches require that at least one text field contains text, and (c) Irrespective of the reported matches of the search query, a maximum of 100,000 are isolated in a search query (so that Refine your results and Analyze Results facilities can only be applied to a maximum of 100,000 documents at a time). As a consequence of these limitations, the task of delineating a subject is subdivided into a series of non-overlapping search sequences that each processes at most 100,000 records and collectively process the entire database. The number of searches depends critically on the type of search. At one extreme it requires only one search sequence to identify all articles for a given year in a SSCI subject; in the search select the SSCI database, specify the year, restrict the document type to articles, and enter 'A* OR B* OR C* OR D* OR E* OR F* OR G* OR H* OR I* OR J* OR K* OR L* OR M* OR N* OR O* OR P* OR Q* OR R* OR S* OR T* OR U* OR V* OR W* OR X* OR Y* OR Z* OR 0* OR 1* OR 2* OR 3* OR 4* OR 5* OR 6* OR 7* OR 8* OR 9*' into the 'SOURCE TITLE:' field. At the other extreme it requires over 500 search sequences to identify all WoS documents in the subject category of 'Information Science & Library Science'. The reason for the much larger number of search sequences is that in the first task only SSCI articles for a single year need to be processed, whereas in the second task all WoS documents for every year need to be processed. For a single year it requires a single search sequence to isolate all SSCI articles, but it can take several search sequences to isolate all SCI articles. The reason is that there are many more SCI articles than SSCI articles.

None of the above limitations apply to Scopus, so that articles in a subject can always be isolated in a single search sequence. Scopus offers the option of specifying the broad subject area on the Basic Search page (Life Sciences, Health Sciences, Physical Sciences, Social Sciences) and the specific subject category on the page that displays the results of searches.

Delineation of subjects conducted using Web of Science subject categories or the JCRs is at the journal level rather than the article level. This results in a number of incongruities: (a) Several articles that have been highly cited in IS&LS were not published in journals classified as IS&LS (e.g., articles by Lotka and Garfield), (b) Some journals such as MIS Quarterly are in the delineation of IS&LS although some might not regard them as LIS journals, and (c) Journals that moved into or out of LIS are included in entirety. Exceptions to (c) can occur due to journal name changes; for instance, all the articles in IRE Transactions on Information Theory are classified as IS&LS, whereas none of the articles in its renamed version IEEE Transactions on Information Theory is classified as IS&LS. Another limitation is that the criteria for inclusion of journals in both the JCR and WoS subject categories may have varied over time; this may apply particularly to WoS articles that were added retrospectively (for instance, the articles prior to 1945 were added in 2005 in the Web of Science Century of Science initiative). Another limitation of the WoS subject categories is the addition of new journals; there were five more journals in IS&LS in September 2007 than in March 2007. None of these had highly cited articles. These limitations can be countered to some extent by examining the extent to which the citing documents are in IS&LS. This provides some indication of the effectiveness of how the articles and their journals have been categorised as IS&LS; in general, higher percentages of citations in IS&LS are more indicative of the effectiveness of this categorisation than lower percentages. Figure 4.2 presents the results of an investigation of this type, but for technical reasons, obtains the percentages in IS&LS WoS ('Information Science & Library Science' delineated on the entire WoS) as opposed to IS&LS.

Figure 4.2 presents, for each of the 77 most highly cited articles in IS&LS (i.e., the most highly cited tenth of a percent) the percentage of citing documents in IS&LS WoS. For more than three quarters of the articles the percentage of citing documents in IS&LS WoS exceeds 20%; this indicates that a high proportion of the articles is likely to be LIS in content. Section 5.3 compares the journals of the 11 articles in Figure 4.1 with fewer than 10% of their citing documents in IS&LS WoS. In Figure 4.2, the vertical axis represents the number of journals whose percentage of citing documents lies in the range presented in the horizontal axis.

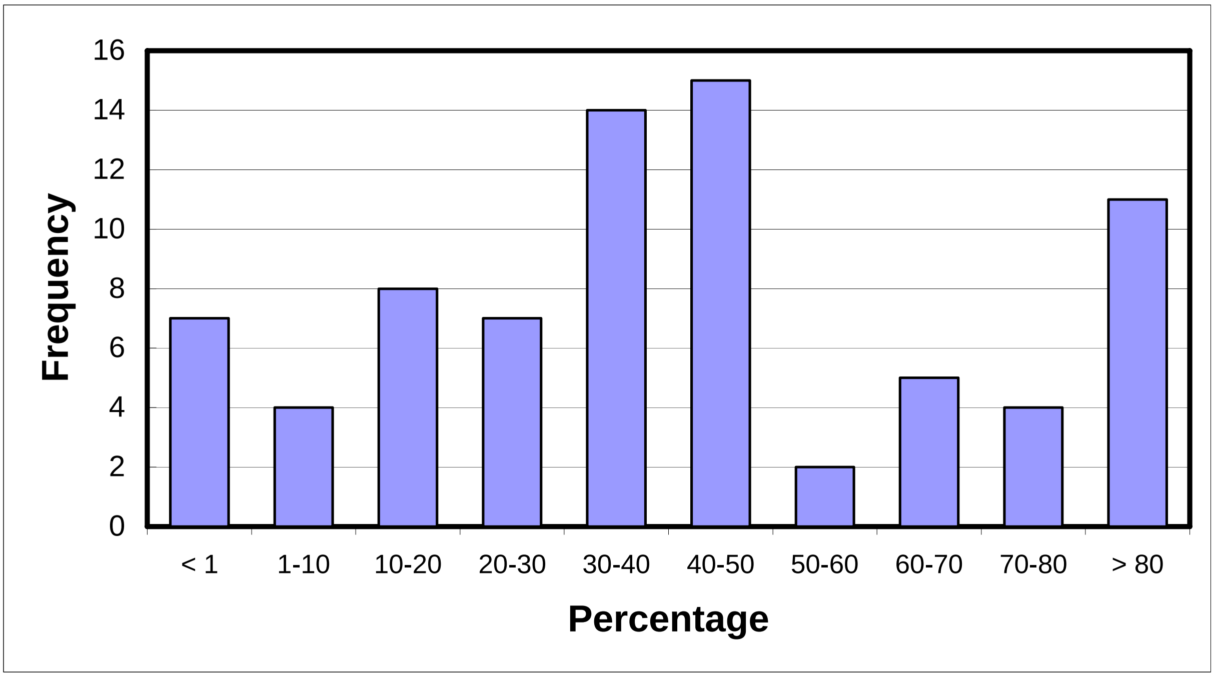

Figure 4.2: The percentage of citing documents in the WoS IS&LS category for the most highly cited 77 articles in IS&LS

4.4: Identifying IS&LS articles by information scientists

Chapter 6 uses wildcards to identify articles by influential information scientists. There are several approaches to identifying articles by individuals. The approach used here is to limit the search to IS&LS articles in SCI and to use wildcards in the AUTHOR' text.

For example, in order to identify articles by Paul B. Kantor using the search, 'Kantor P*' was entered in the 'AUTHOR' text field and 'Refine your results' was used to limit the results to IS&LS articles. The 'Authors' option of 'Refine your results' listed two authors with the name of Kantor (PB and P). The articles by KANTOR, PB were assigned to the influential information scientist Paul B. Kantor. However the articles by KANTOR, P were checked individually for evidence as to whether Paul B. Kantor was the author. This was implemented by examining the WoS article records and conducting Internet searches. Sometimes the article records provided strong evidence of the identity of the author (e.g., the record for 'Using interview data to identify evaluation criteria for interactive, analytical question-answering systems' lists 'Kantor, Paul' as an author); on occasions where there was inconclusive evidence regarding the author's identity (e.g., the record for 'Identification of effective predictive variables for document qualities' provides limited evidence that the influential information scientist P.B. Kantor is an author) searches were conducted on Google and Google Scholar to seek to ascertain the author's identity. The evidence from examining the article records and conducting Internet searches was that all the articles by Kantor P were by Paul B. Kantor.

The search for articles by Donald W. King listed seven authors with the name of King (DW, D, DN, DB, DK, DL and DR). The evidence from examining the article records and conducting Internet searches was that none of the articles by King D was by Donald W King.

4.5: Data processing and Analysis

Although the researcher is able to write programming utilities in several programming languages, programming utilities were not used in the data processing. The rationale is that the processing can readily be done using Microsoft Word and Excel and using these packages avoids the risk of software bugs introducing errors in the analysis. Word was used to sort data (both in tables and free text) and Excel was used to conduct arithmetical calculations, to check the data and to construct graphs.

SPSS was used to conduct the data analysis. The most commonly used option was the Spearman Test, but in addition SPSS was used to **conduct Pearson** Tests, and to obtain a scatter graph.

4.6: Concluding remarks

The methods of data collection presented in this chapter are all used in more than one of the studies presented in Chapters 5 to 8; methods used in a single study are presented in the chapter containing the study. The methods are considered again in the final two chapters in the discussion of the limitations of this study and the areas for further research.

Chapter 5: Characteristics of highly cited LIS articles

Chapter 5 addresses the first global research question: What are the characteristics of the most highly cited articles in Library and Information Science? It consists of five sections: Introduction, Method and Data, Results, Discussion and Concluding remarks. The limitations shared with other chapters are presented in Chapter 9 and the related further research is presented in Chapter 10. Chapter 5 presents the research in Levitt and Thelwall (in press 2008a) that builds on Levitt and Thelwall (2007a).

5.1: Introduction

This study identifies the most highly cited 0.1% of the 77,220 SSCI articles in the subject category of 'Information Science & Library Science' (IS&LS) published prior to 2007. These 77 articles, listed in Table App.1 (Appendix), are used to investigate characteristics of both the highly cited articles and their first authors. The rationale for investigating highly cited articles is that high citation is associated with research quality and consequently, findings on highly cited articles could increase understanding of the quality of research.

This study examines disciplinarity, first authors and citation patterns. One investigation of disciplinarity focuses on the link between multi-disciplinarity and high citation. A reason for examining this topic is that if multi-disciplinary research is cited on average significantly more often than research in a single discipline, it may be worth encouraging multi-disciplinary research. The investigation of first authors examines the citation profiles of the first authors of the highest cited articles. Such profiles provide information that can be used to help identify and allocate resources to those who are more likely to produce highly cited research. The investigation of citation patterns examines the prevalence of late citation amongst the 77 articles. If citation is an indication of research influence, then citation patterns may indicate how this influence has changed over time.

Previous research has addressed the issue of inter-disciplinarity in IS&LS research. Rice and Crawford (1992) identified some possible areas of convergence between the fields of communication and library and information science. Meyer and Spencer (1996) found that library science articles were cited in computer science, medicine, psychology, the social sciences, and general sciences. Tang (2004a) found that Information and Library Science "attracts a significant wide spectrum of disciplines from the domains of science, social science, and the humanities, and that the kinds of disciplines interested in the field vary by year." Other articles on disciplinarity and Information and Library Science include Carlin (2003) and Tang (2004b). Whilst these articles point to considerable disciplinary overlap between Information and Library Science and other disciplines, they do not examine this disciplinary overlap for highly cited articles. This chapter quantifies the disciplinary overlap for a collection of highly cited articles, and compares their overlap with the complete set of articles classified as IS&LS.

In terms of citation profiles, Cronin and Meho (2007) examined the patterns of creative output of renowned information scientists and Cronin and Meho (2006) and Oppenheim (2007) evaluated the h-indexes of influential information scientists. The h-index is defined as the largest number (h) of documents that are cited h or more times (Hirsch, 2005). Several studies, including Hirsch (2005), Batista, Campiteli and Kinouchi (2006), Braun, Glänzel and Schubert (2006) and Van Raan (2006), also use the h-index in various informetric investigations. Whilst previous research has focused on notable researchers in information science it has not examined the citation profiles of the first authors of the most highly cited IS&LS articles. The chapter examines this aspect with a view to identifying whether all the first authors have high h-indexes in IS&LS.

As described in Section 2,4 patterns of annual citation have also been previously researched (e.g., Aversa, 1985; Cano & Lind, 1991; Aksnes, 2003; McCain, 2007). Although previous research has investigated lateness of citation in different subject areas (e.g, Glänzel, Schlemmer, & Thijs, 2003; Van Raan, 2004), apart from Levitt & Thelwall (2007a), it has not examined late citation amongst the most highly cited articles in IS&LS.

Chapter 5 examines this aspect of IS&LS by addressing the following research questions:

1. How does the level of disciplinarity of the most highly cited articles in IS&LS compare with the level of all the articles in IS&LS and to what extent are the disciplinary frequencies of the most highly cited articles mirrored in the frequencies of all the articles in IS&LS?

2. The distribution of the h-indexes of the first authors of the most highly cited articles is examined with a view to establishing whether all first authors have high h-indexes.
3. Define an author's h-index in IS&LS as the h-index of all documents published by the author and classified as IS&LS. How then do the citation profiles of first authors with high h-indexes in IS&LS differ from the citation profiles of first authors with low h-indexes in IS&LS?
4. How widespread is late citation amongst the most highly cited articles?

The rationale for investigating only first authors is that the authors of articles by multiple authors would have had a disproportionate effect on the data. For example, 'Identifying Adverse Drug Events: Development Of A Computer-Based Monitor And Comparison With Chart Review And Stimulated Voluntary Report' has ten authors and the findings would have been skewed if they examined for one article 10 authors and for other articles only 1 author. The Discussion compares the number of articles of the first authors with the number of other authors.

5.2: Method and Data

This chapter investigates the most highly cited tenth of a percent of all articles published prior to 2007 in the subject category of IS&LS. IS&LS is regarded by the JCRs as a social science subject and, as described in Section 3.2, this thesis' investigations of social science subject categories are confined to the SSCI. For this reason, Chapter 5 investigates articles only in the SSCI; as described in

Section 3.2, 93.7% of all WoS articles in IS&LS published prior to 2007 are in the SSCI.

The first research question is addressed by first identifying the disciplinary categories containing the most highly cited IS&LS articles and then comparing for each category the incidence per 100,000 articles of the highly cited articles (Tables 5.1 and 5.2). The second question is addressed by examining the distribution of the h-indexes in IS&LS of the first authors of the highly cited articles (presented in Table App.2, Appendix). The third question is addressed by comparing the citation profiles of first authors with h-indexes in IS&LS over 14 with those with h-indexes in IS&LS of 1 (Table 5.3). The final research question is addressed by examining the number of years between publication and peak citation (Figure 5.1 and Table 5.4).

In order to identify all journals in the SSCI that contain one or more IS&LS articles the database was specified as the SSCI, and a series of non-overlapping search sequences were conducted none of which resulted in more than 100,000 matches, and which collectively covered exactly once every SSCI article published prior to 2007. Each search sequence sought to identify all journals containing IS&LS articles for a specified time period (year or set of years), and the time periods collectively covered exactly once all years prior to 2007. In order to cover every journal, each search sequence specified that the journal name was 'A* OR B* OR C* OR D* OR E* OR F* OR G* OR H* OR I* OR J* OR K* OR L* OR M* OR N* OR O* OR P* OR Q* OR R* OR S* OR T* OR U* OR

V* OR W* OR X* OR Y* OR Z* OR 0* OR 1* OR 2* OR 3* OR 4* OR 5* OR 6* OR 7* OR 8* OR 9*'.

This list of all journals containing at least one IS&LS article obtained from the procedures in the previous paragraph was used to isolate the IS&LS articles. The list was divided into sets containing at most 50 journals and for each set the articles in each set were isolated using Boolean 'OR' searches; the reason why each set was limited to 50 journals is that 50 is the limit for the number of Boolean terms in the Advanced search facility. The articles in IS&LS were isolated by combining the articles in each set (using a Boolean search). Finally the articles in IS&LS were ranked by citation (using the sort by 'Times Cited' facility) and the first 77 articles on the list were the most highly cited 77 articles in IS&LS.

5.3: Results

The method above produced a list of the most highly cited 77 articles out of the 77,220 articles published in IS&LS prior to 2007. The highest cited of these received 995 citations and the lowest 131 citations. The earliest of the 77 articles was published in 1956 and the latest in 2003; the two most frequent decades of publication were the 1990s (44%) and the 1980s (29%). The articles were published most frequently in MIS Quarterly (36%). Over 60% of the articles were by more than one author (26% by precisely two authors and 35% by more than two authors).

For each article the percentages of the citing articles that are in IS&LS were next examined. The rationale is that this provides some indication of the effectiveness of how the articles have been categorised as IS&LS. It seems that, in general, higher percentages of citations in IS&LS are more indicative of the effectiveness of this categorisation than lower percentages. As presented in Figure 4.2, the percentage of citing documents classified as IS&LS (solely or in addition to one or more subjects) ranged from 0.0% to 91.8%.

Within the data set there is an issue of disciplinary classification ambiguity because of the many articles that were rarely cited from IS&LS. Five of the articles with less than 1% of their citing documents in IS&LS were published in IRE Transactions on Information Theory; the journal's only other article, amongst the 77 most highly cited, had 2.1% of its citing documents in IS&LS. One of the articles with less than 1% of their citing documents in IS&LS was the unique article among the 77 most highly cited that was published in International Journal of Geographical Information Science. The remaining article with less than 1% of its citing documents in IS&LS was published in Social Science Information Sur Les Sciences Sociales; the journal's only other articles amongst the 77 most highly cited had 2.7% and 8.7% of their citing documents in IS&LS. The remaining articles with 1 to 10% of their citing documents in IS&LS was published in Knowledge Acquisition; no other article from this publication was amongst the 77 most highly cited. For all SSCI articles in the IRE Transactions On Information Theory published prior to 2007 only 108 of the 4,762 citing documents were in IS&LS (2.3%); the percentage for Knowledge Acquisition was 6.8% (54/796), for Social Science Information

Sur Les Sciences Sociales 11.4% (427/3,760), and for International Journal of Geographical Information Science 11.0% (243/2,216). In summary, the 11 articles with fewer than 10% of citations in IS&LS were published in journals with low percentages of IS&LS citations. Moreover, for the 77 articles, the Spearman correlation between citation ranking and percentage of citing document in IS&LS (not excluding multiple classifications) is .27 (p=.05), indicating that more highly cited articles tend to have a lower percentage of IS&LS citations.

5.3.1: Level of disciplinarity

Here the first research question concerning disciplinarity levels is addressed and the results summarised in Table 5.1. The incidence of the most highly cited articles per 100,000 articles for solely IS&LS is 5.2, and for IS&LS and at least one other subject classification it is much higher at 193. For five of the eight subjects in Table 5.1 the incidence is over 155 and for the remaining three subjects the incidence is 0. The h-indexes on the 7th March 2008 of the sets of articles presented in the table ranged from 9 for Multidisciplinary Sciences to 104 for 'Computer Science, Information Systems' and for 'Solely IS&LS' the h-index was 44.

In Table 5.1, 'IS&LS articles' denotes the number of IS&LS articles, 'Number of articles' the number of most highly cited articles (top 1/10th of a percent) and 'Incidence per 100,000' is the number of most highly cited articles normalised to 100,000 articles.

Table 5.1: Frequencies of IS&LS articles and most highly cited in the eight subjects in which there were the most articles in IS&LS

Subject	IS&LS articles	Most highly cited IS&LS (0.1%)	
		Number of articles	Incidence per 100,000
Computer Science, Information Systems (CSIS)	24,238	65	268
Computer Science, Interdisciplinary Applications (CSIA)	4,060	8	197
Multidisciplinary Sciences	2,614	0	0
Management	2,041	32	1,568
Law	2,565	0	0
Communication	1,871	0	0
Medical Informatics	1,822	7	384
Social Sciences, Interdisciplinary	1,926	3	156
Solely IS&LS	38,261	2	5
IS&LS and at least one other subject	38,959	75	193
All of IS&LS	77,220	77	100

Table 5.2 presents data on the subjects in which there are at least one of the most highly cited articles. The incidences were 5.2 for solely IS&LS and a huge 1,575 for IS&LS, CSIS and Management. The results indicate that inter-disciplinary research is much more likely to be highly cited in IS&LS than research solely within IS&LS. This finding is corroborative evidence that inter-disciplinary research is more likely to be highly cited than research within a single discipline. It is conjectured that inter-disciplinary research in fields for which the average number of citations per article is higher than the norm for Web of Science is even more likely to be highly cited. In Table 5.2, 'IS&LS articles' denotes the number of IS&LS articles, 'Number of articles' the number

of most highly cited articles (top 1/10th of a percent) and 'Incidence per 100,000' is the number of most highly cited articles normalised to 100,000 articles.

Table 5.2: Frequencies of IS&LS articles and most highly cited in subject or subjects in which there is at least one of the most highly cited articles (data from March 2007)

Combinations of subjects	IS&LS articles	Most highly cited IS&LS (0.1%)	
		Number of articles	Incidence per 100,000
Solely IS&LS	38,261	2	5.2
IS&LS and CSIS	20,209	30	148
IS&LS, CSIS and Management	1,787	28	1,567
IS&LS and CSIA	1,823	1	55
IS&LS, CSIS, CSIA and Medical Informatics	1,822	7	384
IS&LS and Social Sciences, Interdisciplinary	1,511	3	199
IS&LS, CSIS, Geography and Geography, Physical	420	1	238
IS&LS and Management	254	4	1,575
IS&LS and Geography	201	1	498

5.3.2: The h-indexes of the first authors

Here the second research question concerning the distribution of the h-indexes of the first authors of the most highly cited articles is addressed. Bates (MJ) was first author of three of the 77 most highly cited articles, and Bates (DW), Belkin, Brancheau, Compeau, Orlikowski, Robertson, Salton, Saracevic and White were each first author of two articles. Table App.2 (Appendix) presents the h-indexes in IS&LS and number of publications in IS&LS of the 66 different first authors of the 77 most highly cited articles. This study examines first authors rather than all authors as examining all authors would have resulted in multiple author articles having a disproportionate impact on the findings.

Specifically, Jha, Kuperman, Teich, et al. (1998) with ten authors and Bates, Teich, Lee, et al. (1999) with eight authors would have been very over-represented. In addition, unlike Chapter 6, it examines the exact name of author (e.g., Bates MJ rather than Bates M*). The rationale is, that for articles published more than two decades ago, there is often not enough information on the Internet to identify which of the articles located using a wildcard in the initial are by the author of the highly cited article, and inaccuracies from using the method would have been particularly prevalent amongst the earliest published articles.

In Table App.2, the h-indexes of first authors vary from 18 to 1. Although 34.8% of first authors had h-indexes greater than 7, 27.3% had h-indexes of 1 or 2. Six researchers, who were first authors of more than one article, had h-indexes greater than 13, and the other 3 first authors of multiple articles had h-indexes of 6 or 7. The 24 authors with h-indexes greater than 7 published 737 articles and these were cited in total 17,357 times (average 23.55). More than 10% of the articles were in four subjects other than IS&LS; 71% of the articles were also classified as Computer Science, Information Systems; 25% as Computer Science, Interdisciplinary Applications; 14% as Medical Informatics and 13% as Management. The 26 authors with h-indexes greater than 7 published 173 articles and these were cited in total 8,365 times (average 48.35). More than 10% of the articles were in two subjects other than IS&LS; 70% of the articles were also classified as Computer Science, Information Systems and 41% as Management. The remaining 22 authors published 56 articles and these were cited in total 6,239 times (average 111.41). More than

10% of the articles were in two subjects other than IS&LS; 71% of the articles were also classified as Computer Science, Information Systems and 38% as Management.

The results indicate considerable variation in the h-indexes of the first authors. That more than 30% of the first authors had h-indexes in IS&LS of 1 or 2 indicates that an author with a low h-index in a subject can produce a highly cited article in the subject. This indicates that the link between high index and high citation has a considerable percentage of exceptions. The high percentage of first authors with an h-index lower than eight (66.7%) may also indicate that IS&LS has benefited considerably from articles published in adjoining fields.

5.3.3: Comparing the citation profiles of first authors with high and low h-indexes

Table 5.3 presents a comparison of the citation profiles of the 9 first authors with an h-index in IS&LS greater than 13 with the profiles of the 10 first authors with an h-index in IS&LS of 1. The final four columns provide a citation profile of documents that cite the IS&LS articles. 'Documents' denotes the total number of citing documents, 'Journals' the number of journals in which they were published, 'Subjects' the number of subject categories in which they were published, and 'IS&LS WoS' the number of citing documents in IS&LS for the entire WoS (for technical reasons, WoS was used rather than the SCI). In order to produce the table the 'Analyse' facility was preferred to the 'Refine your results' facility, as the former lists up to 500 items and the latter lists no more than 100 items.

Table 5.3: Comparing the citation profiles of first authors with high h-indexes in IS&LS with those with h-indexes in IS&LS of one (excluding self-citation by the first author)

Author	h	Articles published	Citing documents			
			Documents	Journals	Subjects	IS&LS WoS only
Salton G	18	41	1,138	361	62	539
Saracevic T	17	67	788	180	49	632
Schubert A	17	65	528	128	78	381
Bates DW	17	44	694	232	70	111
Bates MJ	17	37	680	136	34	570
Benbasat I	17	29	756	169	51	278
Robertson SE	16	38	741	186	58	466
Belkin NJ	16	25	672	143	38	536
Spink A	15	70	450	129	37	313
Iacovou CL	1	2	148	57	27	46
Wrangham RW	1	2	151	43	21	0
Daft RL	1	1	209	80	31	55
Gallager RG	1	1	280	72	19	0
Gruber TR	1	1	612	309	70	34
Kane B	1	1	154	91	45	20
Mata FJ	1	1	129	45	13	45
Max J	1	1	438	119	41	0
Stockwell D	1	1	150	75	35	0

In Table 5.3 the first authors with high h-indexes in IS&LS published on average 46 articles in IS&LS. On average these were cited in 716 citing documents, which were published on average in 185 different journals in 53 different subjects including IS&LS, and 425 of the citing documents were classified as IS&LS WoS. Six of the 10 first authors of more than one highly cited article had an h-index of at least 15 in IS&LS. On average the authors with low h-indexes of 1 in IS&LS published 1.2 articles in IS&LS. On average these were cited in 252 citing documents, published in 99 journals in 34 subjects including IS&LS, and 22 of the citing documents were classified as IS&LS WoS. The most striking difference in the citation profiles is in the percentage of citing documents in IS&LS WoS: For the authors with high h-indexes in IS&LS, in total 59.3% of the citing documents were in IS&LS WoS, whereas for the authors with h-indexes of 1 in IS&LS the percentage was only 8.8%. There is considerable difference between the citation patterns of first authors with high and with low h-indexes in IS&LS and, in particular, articles by authors with high h-indexes were cited much more in IS&LS than those with low h-indexes. Some of the latter are highly cited in other fields, for example, 78% of Joel Max's 444 citations were in 'Engineering, Electrical & Electronic' WoS. Alternatively, some may be "one hit wonders" that have not published a significant body of work, which seems be the case for Mata FJ; in March 2008 the entire WoS held only two publications by Mata FJ; 'Information technology and sustained competitive advantage: A resource-based analysis' was cited 167 times and the other article was cited once.

5.3.4: Late citation

Here the fourth research question concerning the extent of late citation present amongst the most highly cited articles is addressed. Tables 5.4 and 5.5 present the articles for which the peak year of citation was either less than five years or more than forty years after the year of publication.

Table 5.4: Year of publication, year of peak citation and citation ranking for the articles for which the citation peak was less than five years after the publication year

Author(s)	Title	Publi-cation	Peak citat-ion	Citat-ion rank
Kane B, Sands DZ	Guidelines for the clinical use of electronic mail with patients	1998	2001	42
Huber GP	Issues in the design of group decision support systems	1984	1987	55
Saracevic T, Kantor P, Chamis AY, et al.	A study of information seeking and retrieving .1. Background and methodology	1988	1991	60
Brancheau JC, Janz BD, Wetherbe JC	Key issues in information systems management: 1994-95 SIM Delphi results	1996	1999	72
Brancheau JC, Wetherbe JC	Key issues in information-systems management	1987	1991	33
Cimino JJ, Clayton PD, Hripcsak G, et al.	Knowledge-based approaches to the maintenance of a large controlled medical terminology	1994	1998	52
Ingwersen P	Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory	1996	2000	64

Table 5.5: Year of publication, year of peak citation and citation ranking for the articles for which the citation peak was more than forty years after the publication year

Author	Title	Publi-cation	Peak citat-ion	Citat-ion rank
Shannon CE	The zero error capacity of a noisy channel	1956	1998	67
Hu M	Visual-pattern recognition by moment invariants	1962	2005	2
Gallager RG	Low-density parity-check codes	1962	2006	11
Chomsky N	3 models for the description of language	1956	2005	34

In Table 5.4, the 7 articles for which the peak year of citation was less than five years after the year of publication were published on average in 1991.9 (6.8 years later than the average year of publication of the 77 articles). In Tables 5.4 and 5.5 the average citation ranking for the 7 articles with early citation peaks was 54, whereas for the 4 articles with late citation peaks the average citation ranking was 29. This indicates the possibility of a correlation between citation ranking and late citation peak for a set of highly cited articles.

The following Spearman correlations on all 77 articles were evaluated: (a) Correlation between citation ranking and year of publication .18, (b) Correlation between citation ranking and peak year of citation -.36 (p=.01), and (c) Correlation between citation ranking and number of years between peak year and publication year -.46 (p=.01). These correlations are reflected in the findings that for the articles with citation rankings in the top quartile (1 to 19) the average year of publication and peak years were 1981.9 and 2004.1, whereas the corresponding figures for the articles ranked in the bottom quartile (59 to 77) were the years 1988.6 and 1999.2. In addition the percentage of the articles ranked 1 to 23 with citation peak years of 2003 or later was 95.7%, whereas the percentage for the articles ranked 24 to 77 was 50.9%.

Figure 5.1 presents the year of publication and peak citation ranking of the 77 articles.

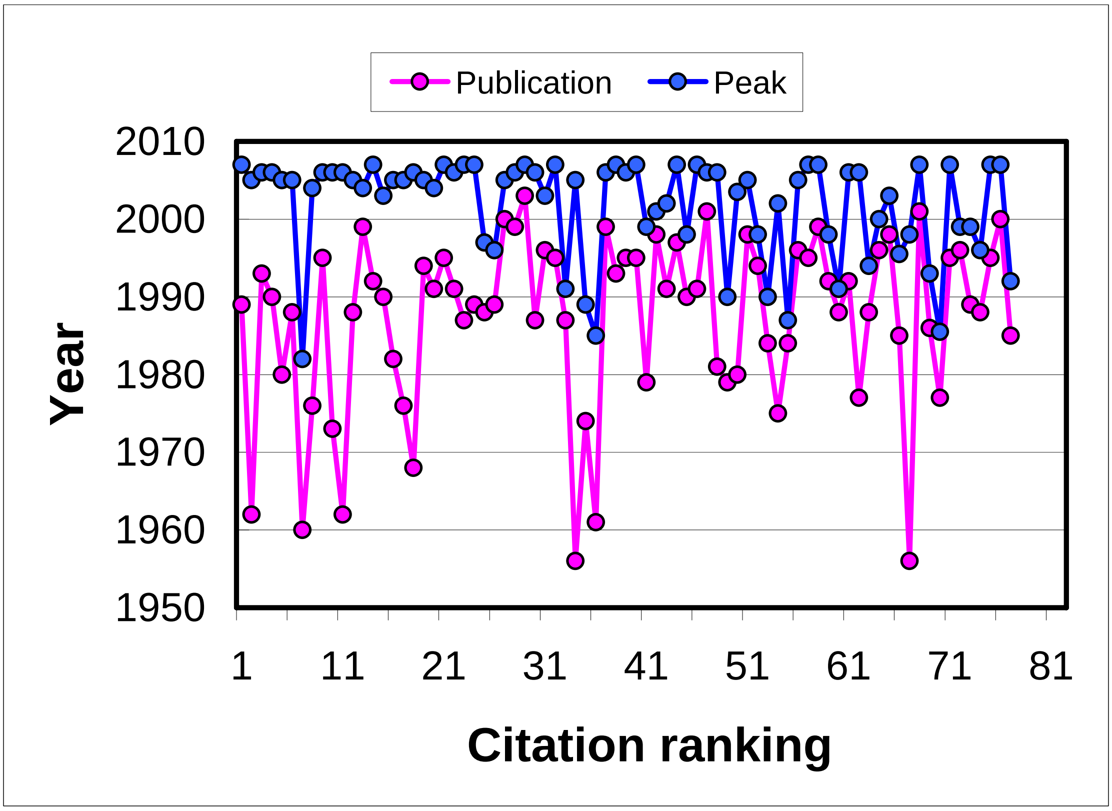

Figure 5.1: Publication year and peak citation year of the 77 most highly cited IS&LS articles

5.4: Discussion

Some interesting findings were made relating to the questions posed in the Introduction. Regarding disciplinarity, the incidence per 100,000 IS&LS of highest cited articles was much lower for articles solely in IS&LS than for those in IS&LS and at least one other subject (9.9 compared with 185.7). It is conjectured that research in IS&LS tends to be incremental (hence individual papers tend not to be highly cited) and that many of the most influential research ideas and methods come from adjacent fields. Hence the promotion of inter-disciplinary research in IS&LS may be conducive to improving the quality of its research.

Regarding the h-indexes of first authors, 27.3% had h-indexes in IS&LS of less than three. Regarding the citation profiles of first authors, a much lower percentage of the citing documents of authors with low h-indexes in IS&LS were classified as IS&LS (for the authors with IS&LS h-indexes exceeding seven in total, 59.3% of the citing documents were in IS&LS, whereas for the authors with IS&LS h-indexes of one the percentage was 7.1%).

Regarding the patterns of annual citation, it is interesting that there is a correlation between citation ranking and lateness of citation and that the twenty most highly cited articles reached their citation peak on average 22 years after publication. This indicates that high quality ideas and methods in IS&LS are often exploited many years after they were published. Although the obliteration phenomenon proposed by Merton (1968) and Garfield (1975, 1993) may have reduced the level of citations of some of the articles examined, many still had late citation peaks.

There are several limitations to this investigation. The incidences and h-indexes in Tables 5.1 and 5.2 that are based on fewer than 5,000 articles need to be interpreted with caution, and the h-index itself is not a reliable indicator on its own (Bornmann & Daniel, 2007). The h-indexes in Table App.2 do not take into account that some individuals are not the first authors of a high percentage of the documents that contribute to their h-index (e.g., Benbasat has an h-index of 17 in IS&LS, but this figure would be 5 if the h-index were evaluated solely on IS&LS documents in which Benbasat is the first author). Furthermore, when comparing the h-indexes of the authors, different authors may have specialised in different subject combinations that have widely differing citation cultures. Although the high frequency of recent citation is interesting, these findings are on a small number of articles and may not be typical of subjects other than IS&LS. It is interesting that for IRE Transactions on Information Theory only 124 of 4,876 citing documents were in IS&LS (2.5%). Limitations shared with other chapters are presented in Chapter 9.

It is interesting to find out whether in general the first author of an article published more articles than the other authors. In order to investigate this question the number of IS&LS articles by the first authors were compared to those for the other authors. In the list of articles in Table App.1, 47 of the articles are by more than one author. In order to obtain a representative sample these articles were ranked in order of decreasing citation and every second article selected. For each article the number of IS&LS articles by the first authors was compared to the average of the number of IS&LS articles by the other authors. A paired t-test established with high significance ($p=.01$) that first authors tended to publish more IS&LS articles than other authors

5.5: Concluding remarks

Some interesting findings were made relating to the questions posed in the Introduction. Regarding disciplinarity, the incidence per 100,000 IS&LS of highest cited articles was much lower for articles solely in IS&LS than for those in IS&LS and at least one other subject (5.2, compared with 193). It is conjectured that research in IS&LS tends to be incremental (hence individual papers tend not to be highly cited) and that many of the most influential research ideas and methods come from adjacent fields. Hence the promotion of inter-disciplinary research in IS&LS may be conducive to improving its quality of research. Nevertheless, it is possible that the high level of citation may not always reflect high quality research and, in particular, may sometimes be partly the result of the inter-disciplinary research inheriting citations from a field that naturally attracts more citations.

Regarding the h-indexes of first authors, the findings indicate that much of high quality research in IS&LS has been produced by first authors who have low h-indexes in IS&LS, either because they are "one hit wonders" (e.g., Mata) or are productive in other fields (e.g., Chomsky, Wrangham). Regarding the patterns of annual citation, the high frequencies of late citation seem to conflict with the low frequencies of late citation found by Glänzel, Schlemmer and Thijs (2003) and van Raan (2004). Two possible reasons for these much higher frequencies are: (a) The criterion used for late citation is less stringent, and (b) The criterion used for highly cited is more stringent.

The first global research question asks what are the characteristics of the most highly cited articles in Library and Information Science. In Chapter 5 the main findings on the most highly cited 0.1% of articles in IS&LS are: (a) The incidence of the articles in IS&LS and another subject (multi-disciplinary) is 37 times that of the articles in IS&LS and no other subject (mono-disciplinary), (b) Sixty-five percent of the first authors of the articles had an h-index in IS&LS of less than eight, and (c) The 19 articles that were cited most (i.e., the most highly cited 25%) reached their citation peak, on average, 23 years after publication. In summary, apparently highly cited IS&LS articles tend to be multi-disciplinary and cited late, but are not necessarily first-authored by influential IS&LS researchers.

The results mentioned above have some corollaries. Finding (a) indicates that many of the most influential research ideas and methods originate in adjacent fields; hence the promotion of inter-disciplinary research in IS&LS may be conducive to improving the quality of its research. Finding (b) indicates that a substantial proportion of influential research in IS&LS has been first-authored by researchers with relatively low h-indexes in the subject; as a consequence it seems important that researchers and policy makers do not assume that influential research can only be first-authored by researchers with high h-indexes in IS&LS. Finding (c) indicates that high quality research in IS&LS is prone to late citation; a likely ramification is that if research quality is judged by citation counts alone, unless a very long citation window is used, research of the highest quality is particularly likely to be under-valued.

Although factors particular to IS&LS (e.g., the diverse educational backgrounds of information scientists and the relative newness of many of the branches of Library and Information) may have contributed to these findings, it seems likely that these findings are not confined to IS&LS. In subjects where the first finding applies, the promotion of inter-disciplinary research may be conducive to improving the quality of research. In subjects where the second finding applies, it seems important for researchers and policy makers not to assume that influential research can only be first- authored by researchers with high h-indexes in the subject. In subjects to which finding (c) applies, if research quality is judged by citation counts alone, unless a very long citation window is used, some very high quality research is likely to be under-valued.

The following findings in this chapter relate to subsequent chapters: (a) Over 60% of the most highly cited IS&LS articles are collaborative, (b) The incidence of the most highly cited articles in multi-disciplinary articles is 37 times higher (193/5.2) than in mono-disciplinary articles, and (c) Late peaks in citation were particularly prevalent amongst the most highly cited quarter of the articles. Collaboration, disciplinarity and late citation are examined more extensively in Chapters 6, 7 and 8.

Chapter 6: Citation levels and collaboration within Library and Information Science

Chapter 6 addresses the second global research question: In general, in Library and Information Science are collaborative articles more highly cited than non-collaborative articles? It consists of five sections: Introduction, Method, Results, Discussion and Concluding remarks. The limitations shared with other chapters are presented in Chapter 9 and the related further research presented in Chapter 10. Chapter 6 presents the research in Levitt and Thelwall (under review). It relates to Chapter 5, which found that over 60% of the most highly cited 77 IS&LS articles were by more than one author.

6.1: Introduction

Over recent years encouragement of collaboration has been a major goal of research policy. Many governments have promoted increased international collaboration in the belief that this results in reduced costs and higher impact research (Katz & Hicks, 1997).

One factor that may have contributed to this belief is the perceived advantage of collaborative research, for instance that collaboration can save time (Fox & Faver, 1984 - in an account of the advantages of collaboration in general) and result in the inclusion of multiple perspectives (Crow, Levine, & Nager, 1992 - an investigation of local collaboration). Another factor is the evidence (described below) that collaboration correlates with high citation. Possibly as a result of government policies, collaboration is increasing in science, albeit unevenly. For example, Gómez, Fernandez, and Sebastian (1999) found that over the period 1991-95 the increase in Latin American international collaboration ranged from 27% (mathematics) to 86% (engineering and technology) and Glänzel (2002) found that over 1980–98 for the SCI the increase in collaboration in general ranged from 17% (mathematics) to 48% (biomedical research).

Investigations of the relationship between citation levels and collaborationhave tended to examine science rather than social science and to use the SCI rather than the SSCI. To partially fill this gap, this chapter examines the relationship between collaboration and citation for the SSCI category of IS&LS.

This research addresses the connection between collaboration and research quality for library and information science through an investigation of connections between citations and collaboration for all IS&LS articles, expanding on Hart's (2007) coverage of two journals. More specifically, the first investigation in this chapter is a longitudinal study of the relationship between levels of citation and collaboration. The second investigation deals with whether the more highly cited articles by IS&LS authors are more highly collaborative than the less highly cited articles, in that it might indicate that collaboration is not only associated with research quality but is conducive to it.

The following specific questions drive this study.

1. Is collaborative IS&LS research more highly cited and how has this changed over time?

2. Are influential IS&LS authors highly collaborative and are the more highly cited articles by influential IS&LS authors more highly collaborative than the less highly cited articles?

6.2: Method

This chapter measures two attributes of collaboration amongst non-anonymous articles: the proportion of articles that are collaborative and the level of collaboration. The first attribute is called by Gómez, Fernandez and Sebastian (1999) 'collaborative rate', and the second attribute is called in this thesis 'collaborative level'. Collaborative levels can be measured in more than one way.

One way of measuring collaborative level, used by Persson, Glänzel and Danell (2004) and Hart (2007), is to compute the number of articles by 1 author, 2 authors, etc. but this is inappropriate here because a single indicator is needed. One such indicator of collaborative level is the average number of authors per article (where 0 is allocated to articles by 1 author, 1 to articles by 2 authors, etc.). Unfortunately, this indicator can be inflated by a small percentage of highly collaborative articles. In order to avoid this problem, an indicator called 'partner score' is introduced to gauge collaborative level. The partner score of an article is defined as 0 if the article is by one author, 1 if by two authors, 2 if by three authors, and 3 if by more than three authors. The average partner score of a set of articles is defined as the average of the partner scores of all articles in the set. From these definitions, all partner scores and average partner scores are between zero and three inclusive.

The decision to allocate the same partner score to all articles by four or more authors is based on data in Table 6.1. The table presents, for all non-anonymous IS&LS articles published in 1995, the frequency of articles for the different numbers of authors. The reasons for allocating the same partner score to articles by more than 3 authors are (a) the small numbers (only 43 or 1.8% of the articles are by 4 authors and 34 or 1.5% by more than 4 authors) and (b) an article by as many as 15 authors would distort the findings (the Discussion presents an example of how the findings would be distorted).

Table 6.1: Frequency of articles for different levels of co-authorship (non-anonymous IS&LS 1995)

Number of authors	IS&LS articles	Number of authors	IS&LS articles
1	1,649	7	2
2	446	9	3
3	165	10	1
4	43	11	1
5	19	15	1
6	7	8, 12 to 14, > 15	0

Using WoS data, this study implements two investigations, one for each research question.

The first investigation compares, for every second year from 1976 to 2004, the average partner scores of the most highly cited 1% of the articles in the year, the articles with citation rankings in the top 1-2%, 10%-12% and 20%-22%, and the un-cited articles. The reason for examining the articles ranked 10%-12% and 20%-22% rather than 2%-10% and 10-20% is that a pilot investigation of IS&LS articles published in 1995 found limited variation in the average partner score in the interval 5%-21% (for the articles in 5.1%-10% it was .87, articles in 10-14.9% it was .80 and articles in 16.9 to 21% it was .78). In view of these limited changes it was decided that changes could be identified more clearly by having gaps between the citation ranges.

In order to obtain more accurate results the calculations are based on as large a sample as possible; for example, average partner scores for un-cited articles are calculated on all un-cited articles. In addition to basing the findings on a large sample this method avoids possible biases due to the way the articles are selected. For instance, articles selected from a list produced by the sort by 'Times Cited' facility can be biased towards a highly collaborative journal. Specifically, for the Journal of the American Medical Informatics Association (JAMIA) for all articles published in even years between 1976 and 2004 the collaborative rate is .91 and average partner score is 1.97; 76% of the first 45 un-cited articles on the list for 2000 are in JAMIA, whereas only 18% of all the un-cited IS&LS articles published in 2000 were in JAMIA.

The second investigation compares the average partner scores of all the IS&LS articles for a selection of influential IS&LS authors. The selected authors were either (a) winners of ASIST's Research in Information Science Award, (b) first authors of very highly cited articles in library and information science (amongst the most highly cited 0.1% of all WoS IS&LS articles), or (c) winners of the Derek John de Solla Price Medal. The use of multiple criteria in this way allows a broader conceptualisation of influential researchers than allowed by any single criterion. The Research in Information Science Award "recognizes an individual or organization for outstanding research contributions in the field of information science" (http://www.asis.org/awards/research.html); the Derek John de Solla Price Medal is awarded "to scientists with outstanding contributions to the fields of quantitative studies of science" (http://www.issi-society.info/price.html), and is included as a recognised award to compensate for any possible U.S. bias in the ASIST award. Recent studies have examined citation profiles of ASIST's winners of the Award of Merit and Research in Information Science Award (Cronin & Meho, 2007), Derek John de Solla Price Medal winners (Egghe, 2006), and Chapter 5 compares the citation profiles of the first authors of the 77 most highly cited IS&LS articles.

The IS&LS articles for each of the selected authors were identified by using a wildcard search in the author's first name. For example, in order to identify articles by Paul B. Kantor, IS&LS articles by 'Kantor P*' were identified; decisions as to whether the author was Paul B. Kantor were made on the basis of WoS article records and Internet searches.

For technical reasons, in both investigations the citation data is the total number of citations from publication to data collection. The extent to which the findings might be affected by this decision to use an open citation window is considered in the Discussion.

6.3: Results

6.3.1: Longitudinal analysis of selected IS&LS articles

This section investigates whether collaborative IS&LS research is highly cited and how this has changed over time. The collaborative rates and average partner scores were calculated for every even year from 1976 to 2004 for articles in the five citation level sets described in the Methods and the results are in Tables 6.2 to 6.5 and Figures 6.1 and 6.2.

Table 6.2, 6.3 and 6.4 present the range of citation, collaborative rates and average partner scores of diverse citation levels and Table 6.5 presents the number of articles on which the range of citation, collaborative rates and average partner scores were calculated.

In Table 6.2, 'Top 1%', 'Top 1 to 2%', 'Top 10 to 12%' and 'Top 20 to 22%' denote respectively the range of citation of the articles among the most highly cited 1%, the most highly cited 1 to 2%, the most highly cited 10 to 12%, and the most highly cited 20 to 22%.

Table 6.2: IS&LS citation ranges over time for four citation levels

Year	Top 1%	Top 1 to 2%	Top 10 to 12%	Top 20 to 22%
1976	36 to 383	18 to 35	5 to 6	3
1978	26 to 74	19 to 25	5 to 6	3
1980	35 to 572	23 to 33	5 to 6	3
1982	27 to 258	20 to 26	6 to 7	3 to 4
1984	34 to 151	21 to 32	6 to 7	4
1986	48 to 135	30 to 48	7 to 8	4
1988	73 to 484	29 to 71	7 to 8	3 to 4
1990	33 to 628	20 to 33	5 to 7	3
1992	41 to 251	25 to 40	8 to 9	4
1994	49 to 246	33 to 47	8 to 10	3 to 4
1996	50 to 191	32 to 50	9 to 10	4 to 5
1998	44 to 156	31 to 42	10 to 12	5
2000	37 to 202	29 to 36	10 to 12	5
2002	39 to 93	26 to 39	8 to 9	4 to 5
2004	18 to 88	15 to 18	6	3

In Table 6.3, 'Top 1%', 'Top 1 to 2%', 'Top 10 to 12%', 'Top 20 to 22%' and 'Uncited' denote respectively the collaborative rate of the articles among the most highly cited 1%, the most highly cited 1 to 2%, the most highly cited 10 to 12%, and the most highly cited 20 to 22%, and the uncited.

Table 6.3: IS&LS collaborative rates over time for five citation levels

Year	Top 1%	Top 1 to 2%	Top 10 to 12%	Top 20 to 22%	Uncited
1976	.33	.42	.17	.31	.17
1978	.39	.30	.40	.19	.19
1980	.24	.24	.26	.18	.22
1982	.41	.18	.35	.27	.21
1984	.41	.28	.30	.23	.20
1986	.53	.26	.20	.25	.21
1988	.94	.50	.42	.31	.22
1990	.50	.63	.36	.26	.17
1992	.58	.41	.48	.33	.21
1994	.73	.56	.49	.43	.30
1996	.65	.64	.55	.52	.20
1998	.91	.45	.56	.52	.17
2000	.87	.82	.57	.39	.19
2002	.84	.80	.73	.57	.22
2004	.67	.75	.60	.66	.21
Mean	**.60**	**.48**	**.43**	**.36**	**.21**

In Table 6.4, 'Top 1%', 'Top 1 to 2%', 'Top 10 to 12%', 'Top 20 to 22%' and 'Uncited' denote respectively the average partner score of the articles among the most highly cited 1%, the most highly cited 1 to 2%, the most highly cited 10 to 12%, and the most highly cited 20 to 22%, and the uncited.

Table 6.4: IS&LS average partner scores over time for five citation levels

Year	**Top 1%**	**Top 1 to 2%**	**Top 10 to 12%**	**Top 20 to 22%**	**Uncited**
1976	.40	.62	.17	.38	.21
1978	.44	.41	.50	.26	.27
1980	.35	.29	.29	.23	.31
1982	.71	.18	.41	.39	.30
1984	.59	.35	.46	.31	.25
1986	.58	.26	.24	.30	.28
1988	1.39	.67	.67	.42	.29
1990	1.03	.91	.46	.36	.25
1992	1.00	.65	.67	.41	.31
1994	1.59	.92	.78	.62	.52
1996	1.26	1.05	.82	.83	.31
1998	1.74	.77	.98	.77	.21
2000	1.57	1.09	.85	.69	.27
2002	1.34	1.43	1.22	.93	.31
2004	1.57	1.65	1.03	1.11	.30
Mean	**1.04**	**.75**	**.64**	**.53**	**.29**

In Table 6.5, 'Top 1%', 'Top 1 to 2%', 'Top 10 to 12%', 'Top 20 to 22%' and 'Uncited' denote the number of articles used to calculate the collaborative rate and average partner score for respectively most highly cited 1%, the most highly cited 1 to 2%, the most highly cited 10 to 12%, and the most highly cited 20 to 22%, and the uncited. In addition, for each year 'Articles in the year' denotes the number of non-anonymous IS&LS articles.

Table 6.5: Number of IS&LS articles on which the collaborative rates and average partner scores were calculated for five citation levels

Year	Top 1%	Top 1 to 2%	Top 10 to 12%	Top 20 to 22%	Uncited	Articles in the year
1976	15	19	64	98	658	1,480
1978	18	20	77	123	830	1,794
1980	17	17	83	100	698	1,723
1982	17	17	69	228	618	1,651
1984	17	21	74	94	613	1,698
1986	19	18	68	133	574	1,746
1988	18	18	62	199	736	1,793
1990	20	25	137	132	829	1,917
1992	19	21	63	73	755	1,896
1994	22	23	79	200	1,056	2,205
1996	23	24	53	184	995	2,246
1998	23	22	73	65	948	2,234
2000	23	22	63	75	981	2,245
2002	23	23	73	187	938	2,219
2004	27	23	62	141	874	2,070
All	**301**	**313**	**1,100**	**2,032**	**12,103**	**28,917**

The summary in Table 6.6 (derived from analysing collaboration in 15,849 articles) indicates that IS&LS articles with more citations on average are more collaborative and have more partners than articles with fewer citations. The table also indicates that the divergence in collaboration is particularly large when comparing the lowest and the highest levels of citation. Specifically, although the collaborative rate and average partner score of 'Top 1%' were respectively 186% and 259% higher than those of 'Uncited', the collaborative rate and average partner score of 'Top 1-2%' were only 33% and 42% higher than those of 'Top 20%-22%'.

Table 6.6: Mean collaborative rates, average partner scores and partner scores (IS&LS articles, every even year from 1976 to 2004)

Citation level	Collaborative rate	Average partner score	Partner score			
			0	1	2	>= 3
Top 1%	.60	1.04	111	87	52	42
Top 1-2%	.48	.75	141	87	33	23
Top 10%-12%	.43	.64	285	218	44	30
Top 20%-22%	.36	.53	378	143	37	21
Uncited	.21	.29	9,171	1,801	614	517

A Pearson Chi Square test on the partner scores derived from Table 6.6 gave 35.000 (p = .09). The high p value indicates that the test found no evidence of a relationship between citation level and partner score.

Figures 6.1 and 6.2 illustrate longitudinal changes in collaborative rates and average partner scores. In the figures for all citation levels apart from un-cited the collaborative levels and average partner scores have increased steadily over the period 1975 to 2004. The oscillations in collaborative rate and partner score in the figures for the top 1% and 1%-2% could be due to the small numbers in these ranges.

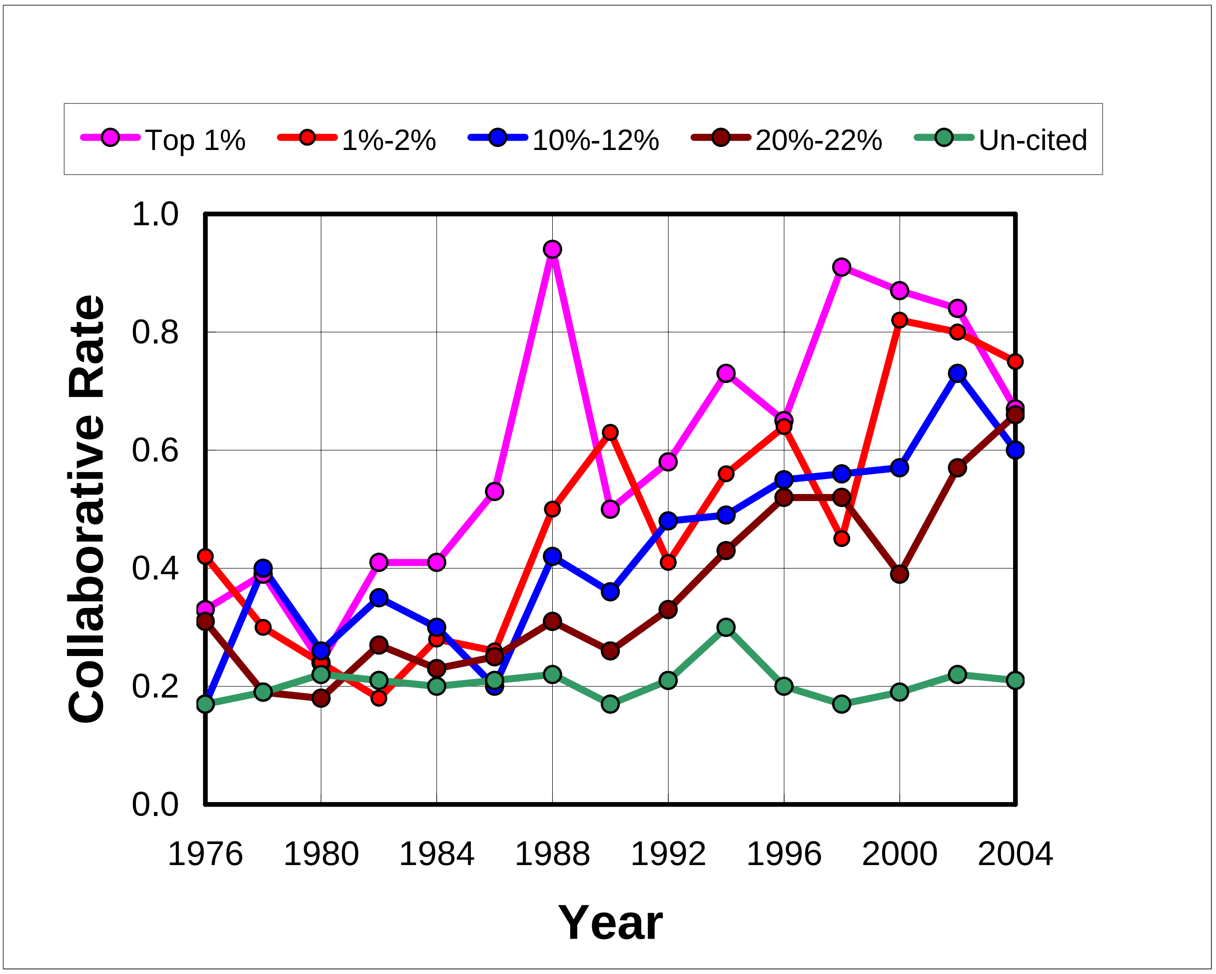

Figure 6.1: Proportion of collaborative articles (Collaborative rates) over time for five citation strata

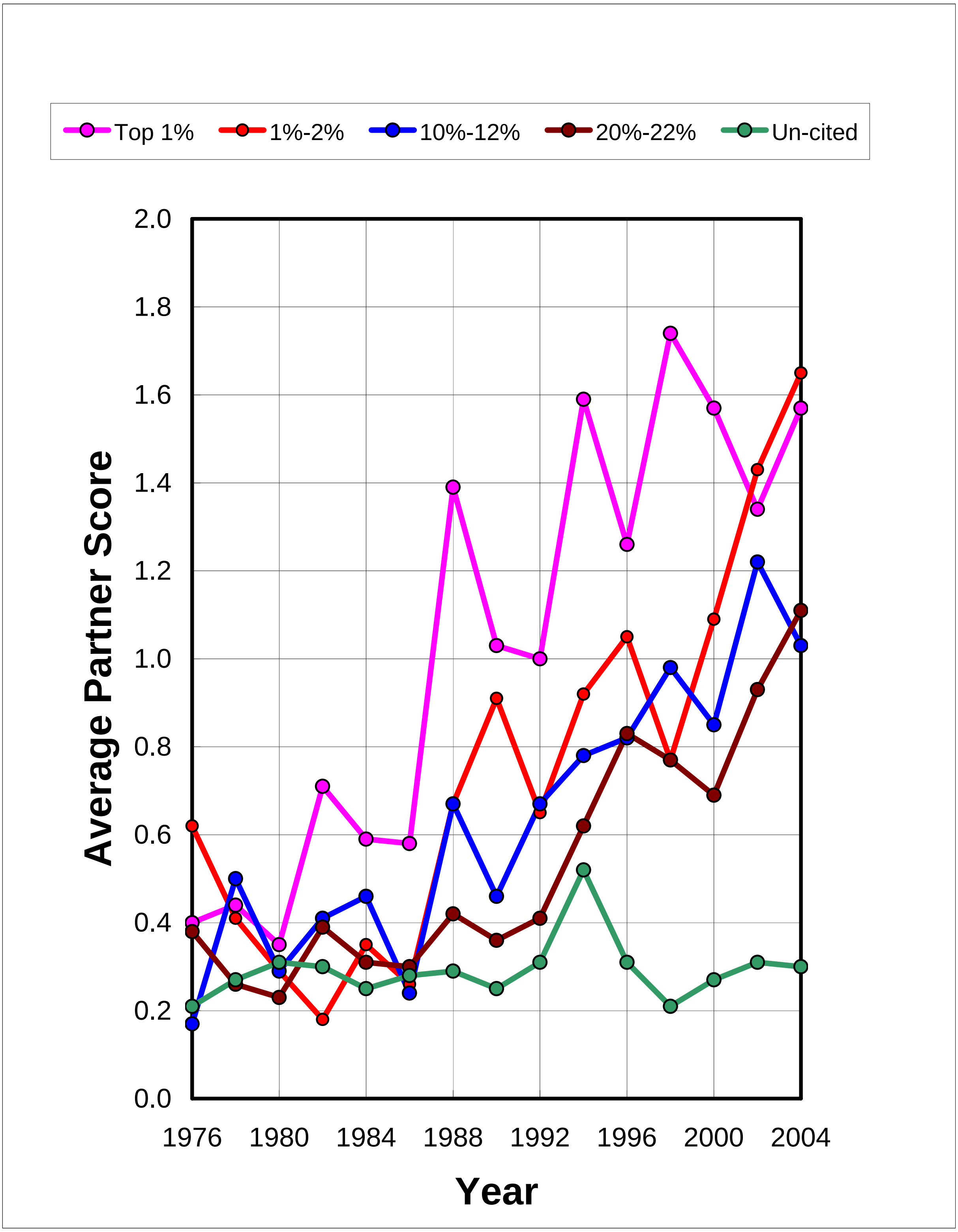

Figure 6.2: Average partner score over time for five citation strata

The upward trends in Figures 6.1 and 6.2 were confirmed using the Spearman test, with the findings presented in Tables 6.7, 6.8 and 6.9. Although the Kolmogorov-Smirnov Normality tests indicate that the Pearson test could be used instead of the Spearman test (throughout p > .05), the Spearman test was used because the correlation of rankings seems more relevant to this study than the correlation of values. In Table 6.7, for all levels of citation, apart from zero citation, the correlations with time (years) are statistically significant (p <.01). In Tables 6.8 and 6.9, for both collaborative rates and average partner score the correlation between every pair of citation levels that does not include zero citation is statistically significant (p <.01), and for none of the pairs that include zero citation is the correlation significant. Moreover, in every case average partner score correlates more strongly than does collaborative rate.

Table 6.7: Spearman correlations between collaborative rates and year and between average partner scores and year for five citation levels

Citation level	Collaborative rate	Average partner score
Top 1%	.80	.85
Top 1-2%	.80	.86
Top 10%-12%	.91	.90
Top 20%-22%	.86	.88
Un-cited	.14	.28

Table 6.8: Spearman correlations in collaborative rate over time for pairs of articles with different citation levels

Citation level	Top 1%	1-2%	10-12%	20-22%	Un-cited
Top 1%	1.00	.67	.77	.72	.19
1-2%	.67	1.00	.79	.76	-.08
10-12%	.77	.79	1.00	.83	.14
20-22%	.72	.76	.83	1.00	.09
Un-cited	.19	-.08	.14	.09	1.00

Table 6.9: Spearman correlations in average partner scores over time for pairs of articles with different citation levels

Citation level	Top 1%	1-2%	10-12%	20-22%	Un-cited
Top 1%	1.00	.74	.83	.82	.13
1-2%	.74	1.00	.85	.84	.20
10-12%	.83	.85	1.00	.87	.27
20-22%	.82	.84	.87	1.00	.26
Un-cited	.13	.20	.27	.26	1.00

6.3.2: Influential IS&LS authors

This section investigates whether influential IS&LS authors are highly collaborative and whether the more highly cited articles by influential IS&LS authors are more highly collaborative than the less highly cited articles.

Table 6.10 presents the average partner score of 35 influential information scientists. Retired and deceased information scientists are included in the study, but information scientists were omitted when it was considered that there were too few articles on which to reliably evaluate the average partner scores. The effect on the findings is that they are based on as large as possible a sample for which there is reliable data. The average year of publication varies between information scientists, but the year of publication is taken into account in Table 6.11. Although 74% of the information scientists have average partner scores between .67 and 1.71, the average partner score ranges from .10 to 2.91 and Marchionini's score is 5.2 times Tenopir's score.

Table 6.10: Average partner score per article for 35 influential information scientists for IS&LS articles published between 1975 and 2004

Information scientist	Score	Information scientist	Score
Bates, DW	2.91	Kuhlthau, CC	.94
Bates, MJ	.54	Leydesdorff, L	.37
Belkin, NJ	1.54	Marchionini, G	1.65
Benbasat, I	1.35	Mccain, KW	.77
Bookstein, A	.74	Meadow, CT	1.15
Braun, T	1.64	Moed, HF	1.71
Brookes, BC	.10	Narin, F	1.17
Cimino, JJ	2.22	Robertson, SE	1.23
Croft, WB	1.10	Salton, G	1.33
Egghe, L	.50	Saracevic, T	.92
Fidel, R	.54	Schubert, A	1.63
Garfield, E	.69	Small, HG	.43
Glanze,l W	1.39	Spink, A	1.36
Griffith, BC	1.00	Straub, DW	1.54
Harter, SP	.67	Tenopir, C	.32
Ingwersen, P	.91	Van Raan, AFJ	1.20
Kantor, PB	1.07	White, HD	.78
Kraft, DH	1.39	**Mean**	**1.11**

The mean partner score of the 35 information scientists in Table 6.10 is 1.11. Table 6.11 lists for each information scientist the average partner scores of highly cited articles (at least 30 citations) with those of low cited articles (5 or less citations). The mean partner score for the highly cited articles is 1.12 and for the low cited articles is slightly higher at 1.15. However, in Table 6.11 the mean year of publication for the highly cited articles is 4.5 years earlier than that for the low cited articles, which makes the difference even less significant. Hence it seems that there is no real difference in collaborative level between highly cited and un-cited articles of the top information scientists. In Table 6.11, 'Partner score' and 'Average year' denote the average partner score and average year of publication of the articles.

Table 6.11: Average partner score and average year of publication for high and low cited IS&LS articles (1975 to 2004)

Information scientist	30 or more citations		5 or less citations	
	Partner score	Average year	Partner score	Average year
Bates, DW	3.00	1999.3	2.83	1999.6
Bates, MJ	.40	1985.1	1.25	1996.0
Belkin, NJ	1.36	1984.5	1.71	1993.3
Benbasat, I	1.31	1995.7	1.25	2001.3
Bookstein, A	.33	1979.3	.85	1990.6
Braun, T	1.67	1990.5	1.18	1994.4
Brookes, BC	.00	1979.6	.00	1980.0
Cimino, JJ	2.80	1994.4	2.15	1999.0
Croft, WB	1.00	1983.5	1.50	1991.5
Egghe, L	.25	1991.8	.45	1998.4
Fidel, R	.50	1988.8	.40	1988.9
Garfield, E	.33	1981.3	1.00	1997.4
Glanze,l W	1.33	1992.9	1.41	1997.2
Griffith, BC	1.50	1980.0	1.00	1988.0
Harter, SP	.33	1991.0	.89	1988.0
Ingwersen, P	.80	1993.9	1.00	1996.8
Kantor, PB	1.60	1987.0	1.11	1991.0
Kraft, DH	1.00	1979.8	1.75	1984.4
Kuhlthau, CC	.60	1990.0	1.43	1995.7
Leydesdorff, L	.00	1992.0	.54	1995.3
Marchionini, G	.75	1990.3	1.80	1998.9
Mccain, KW	.67	1991.5	1.17	1994.4
Meadow, CT	1.00	1980.5	.94	1987.2
Moed, HF	1.88	1993.6	1.25	1996.8
Narin, F	1.30	1983.2	1.40	1996.1
Robertson, SE	1.00	1984.4	1.31	1986.5
Salton, G	1.75	1986.6	.50	1984.6
Saracevic, T	1.63	1993.9	.81	1993.7
Schubert, A	1.40	1987.2	1.47	1992.4
Small, HG	.75	1984.9	.50	1999.2
Spink, A	2.00	1999.0	1.19	1992.8
Straub, DW	1.10	1997.3	1.88	1993.6
Tenopir, C	1.00	1993.7	.25	1994.6
Van Raan, AFJ	2.00	1992.8	1.18	1989.2
White, HD	1.00	1987.0	.82	1996.1
Mean	**1.12**	**1988.8**	**1.15**	**1993.3**

The mean values of the average partner score for '30 or more citations' is very close to that for '5 or less citations'. In order to check whether there is a statically significant correlation between the average partner score of the high and low cited articles, a paired t-test was conducted. A significant correlation of .60 was found ($p < .01$).

6.4: Discussion

As described in the Method, for technical reasons this paper calculates citations to date rather than using a fixed citation window. One could conjecture that one impact of not using a fixed citation window is that in general older articles would have more citations than recent articles (as they have more years in which to be cited). In order to test this conjecture data on the evolution of citation with time was gathered and presented in Figure 6.3. The figure presents the total number of citations to date for the IS&LS articles with citation rankings of 1%, 2%, 10% and 20% (called 'First', 'Second', 'Tenth' and 'Twentieth'), as percentages of their 1976 values (36 for 1%, 18 for 2%, 6 for 10% and 3 for 20%).

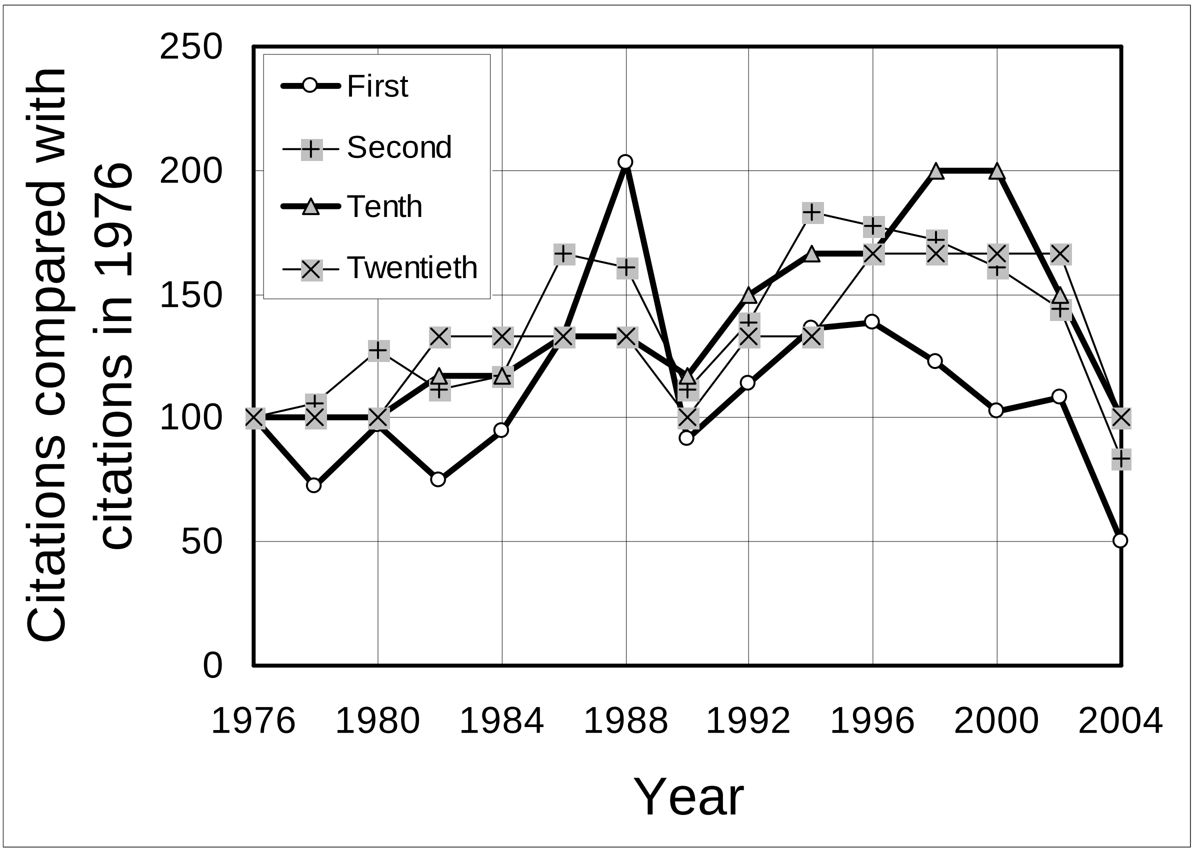

Figure 6.3: Citations to date of the article at the first, second, tenth and twentieth percentile expressed as a percent of the corresponding 1976 values (IS&LS 1976 to 2004)

Figure 6.3 does not suggest that older articles tend to be more heavily cited than more recent articles. Although articles from 2004 are mostly at a disadvantage compared to other years, all groups of articles from 2002 attracted more citations than those from 1976. This suggests that, except possibly for 2004 articles, the citation window used does not significantly disadvantage the most recent articles.

In the Methods, one of the reasons for using the average partner score rather than total number of authors to indicate collaborative level is that large numbers of authors can distort the findings. For example, counting every partner would have resulted in the average partner score for the most highly cited 1%-2% of the articles being much higher in 1994 than in 1996 (1.97 compared with 1.07); the 1994 score would have been distorted by one article by 16 authors and one by 15 authors.

As described in the Method, this Chapter investigates the intervals 10%-12% and 20%-22% rather than 2%-10%, and 10%-20%. The advantage of using citation ranges separated by gaps is that discrete ranges identify gradual change more easily; in order to offset the consequent disadvantage of smaller numbers as large a sample of articles as possible was used.

In common with several studies described in the Related Research the longitudinal investigation found a positive link between collaboration and citation level; for instance, on average the collaborative rate and partner scores for articles cited in the top 1% was 2.9 and 3.6 times those of un-cited articles (Table 6.6). This study also found that the trend towards multiple authorship in biomedical research and chemistry (Glänzel, 2002) was present in IS&LS at all citation levels apart from un-cited. A surprising finding is the difference in behaviour between un-cited articles and the other sets of articles examined: For all citation levels, apart from un-cited, there were strong statistical correlations both longitudinally and between level. In contrast, Hart (2007) did not find a correlation between collaboration and citation in library science, but Hart's investigation was confined to two journals.

The results for the most influential scholars paint a rather different picture. There were significantly different patterns of collaboration. For example, the collaborative levels of Marchionini and Belkin were respectively 5.2 and 4.8 times the collaborative level of Tenopir. This indicates that some influential information scientists have chosen to collaborate much more often than others. The collaborative level of Tenopir is only 3% higher than the average for zero citations, indicating that a high level of collaboration has not been a pre-requisite to become an influential information scientist. Part of the reason for the findings may be due to the sub-field factors; researchers oriented towards computer science (e.g., Marchionini, Belkin) are possibly more likely to collaborate, whereas researchers oriented towards the social sciences/humanities end of IS&LS, such as information behaviour research (e.g., Tenopir, MJ Bates, Fidel), may tend towards the 'lone scholar' mode of research. Of course, all these recognised researchers have mature bodies of work and hence reflect past IS&LS routes to success rather than the achievements of less recognised scholars that are today working towards gaining recognition.

An unexpected finding was that the more highly cited articles by influential IS&LS authors on average are not more highly collaborative than the less highly cited articles; however this finding may be partly explained by the fact that the more highly cited articles were published earlier and that, apart from the un-cited, earlier articles were found to be less collaborative.

Although the results give clear answers to the research question, there are some limitations. First, the findings are dependant upon the WoS selection and categorisation of journals for IS&LS. In particular, some of the IS&LS journals are likely to be seen as outside the library and information science discipline, such as those dealing with information systems and medical informatics. Second, WoS designates subjects at the journal level in that, for a given journal, all articles are given the same subject designation. This designation results in a coarse-grained definition of Library and Information Science; articles are in LIS if, and only if, they are in journals designated to the IS&LS subject category.

As mentioned in Section 2.2, Bordons, Gómez, Fernandez, Zulueta, and Mendez (1996) classify collaborative articles into types that reflect the geographical proximity of the collaborators, but does the link between collaboration and citation depend on proximity? An investigation of the IS&LS articles published in 1995 did not obtain a statistically significant correction between proximity and highly cited collaborative articles. The 100 most highly cited collaborative articles were classified into to six levels of proximity. Table 6.12 presents for each proximity level the number of articles, mean citation ranking and mean number of citations. In Table 6.12, 'Articles' denotes the number of articles with authors at the level of proximity, 'Mean citation ranking' the average of citation ranking of the articles and 'Mean citation' the average number of citations of the articles.

Table 6.12: Number of articles, mean citation ranking and mean number of citation for the diverse levels of proximity (100 most highly cited collaborative IS&LS articles published in 1995)

Proximity level	Articles	Mean citation ranking	Mean citation
Same department	33	53.52	56.27
Same institution	12	59.92	32.17
Same city	8	54.63	52.88
Same state	19	36.05	64.84
Same country	17	54.94	54.18
Different country	11	46.27	45.36

In Table 6.12, 'Same department' denotes that the authors were listed by WoS as being in the same departments and in the same institution, 'Same institution' denotes that the authors were listed in the same institution but not all in the same department, and the others are defined similarly. 'Different country' denotes that not all authors were listed in the same country. For the 100 articles, the Spearman correlation between proximity and citation ranking was less than .1 (p=.34).

6.5: Concluding remarks

Returning to the research question 1, collaboration is clearly associated with higher citation (contrary to Hart, 2007). Moreover, the collaborative rates and levels of the highest four citation strata increased in unison over time, whereas the collaborative rates and levels of the un-cited articles remained low and stable. From this, it seems reasonable to hypothesise that, in general, collaborative research is becoming increasingly significant and influential in library and information science. An implication for authorship is that it may have become increasingly difficult for a sole author to produce highly cited IS&LS research.

The finding in Table 6.6 that the variation in collaboration with citation is particularly large at the lowest and highest citation level is interesting in two respects: (a) It provides a fine-grained picture of the relationship between collaboration and citation level, and (b) Because un-cited articles are very common (from Table 6.5, the non-anonymous un-cited articles form between 32.9% and 47.9% of all articles in the year), it seems important that the particularly low collaborative rate and level of un-cited articles are taken into account when investigating collaboration.

The study on influential information scientists found that although on average they had high collaborative levels, the collaborative level for some authors was less than a fifth of that of other authors. This indicates that although collaboration is associated with research quality it is not a requirement. This was confirmed by the evidence that the level of collaboration of an author's individual articles did not associate with high citation. The more highly cited articles by influential IS&LS authors on average are not more highly collaborative than their less highly cited articles, but in general they were published earlier.

The second global research question asks whether, in general, collaborative Library and Information Science articles are more highly cited than non-collaborative articles. In Chapter 6 the main findings are: (a) Comparing the proportion of collaboration in five strata of citation levels, collaboration in each of the highest four citation strata (all in the most highly cited 22%) increased in unison with time, whereas collaboration in the lowest citation strata (uncited articles) remained low and stable, and (b) The more highly cited articles by influential information scientists, on average, were not more highly collaborative than the less highly cited articles. In summary, highly cited research is particularly collaborative at the general level it but not at the level of individual influential researchers. Moreover, the level of proximity of collaborators does not seem to associate with the citations of the resulting articles (Table 6.12).

Finding (a) indicates that longitudinal changes in collaborative behaviour in IS&LS are not necessarily the same for every citation stratum; hence it seems important when investigating collaboration to examine the behaviour of different citation strata, particularly for uncited articles (uncited IS&LS form about 40% of all IS&LS articles). Although the finding (b) was unexpected, it may be partly explained by the fact that the highly cited articles were published on average 4.5 years earlier than the low cited articles; as a consequence it is recommended, when comparing the collaborative levels of researchers, that differences in the years of publication are taken into account.

Chapter 7: Multi-disciplinarity in science and social science

Chapter 7 addresses the third global research question: In general, are multi-disciplinary articles more highly cited than mono-disciplinary articles? It consists of five sections: Introduction, Method, Results, Discussion and Concluding remarks. The limitations shared with other chapters are presented in Chapter 9 and the related further research presented in Chapter 10. Chapter 7 presents the research in Levitt and Thelwall (in press 2008c) and uses the Normalised Hirsch Index indicator introduced in Levitt and Thelwall (2007b). It relates to Chapter 5, which found that 75 of the most highly cited 77 IS&LS articles to be multi-disciplinary, although only 50.5% of all IS&LS articles are multi-disciplinary.

7.1: Introduction

The beliefs that inter-disciplinary collaboration is conducive to quality in research and that some problems are too complex to be solved in a single discipline underlie the recent policy goal of encouraging collaboration between researchers in different disciplines, especially as part of modern applied inter-disciplinary 'Mode 2' research (Gibbons, Limoges, Nowotny, Schwartzman, Scott, & Trow, 1984). One perceived advantage of Mode 2 research is that it opens knowledge production to a wide range of influences (Leydesdorff & Etzkowitz, 2001).

Inter-disciplinarity has been encouraged in science policy both by creating multi-disciplinary centres and units and by funding multi-disciplinary research projects (Bordons, Zulueta, Romero, & Barrigon, 1999). Many science policy documents express high expectations of the benefits of inter-disciplinary research (Rinia, Van Leeuwen, Bruins, Van Vuren, & Van Raan, 2002). There has recently been a sharp rise in the number of policies and the amount of funding aimed at promoting cross-disciplinary collaboration between different fields, leading to claims that cross-disciplinarity has become the 'mantra of science policy' since the mid 1990s (Rafols & Meyer, 2007). In addition, an entire chapter of Moed (2005) discusses an example of a national Research Council seeking to stimulate trans-disciplinary research.

The purpose of this chapter is to examine the extent to which the level of disciplinarity correlates with citation. Specifically, for diverse subjects in science and social science, it compares the level of citation of the journals classified in more than one subject with the level of citation of the journals classified in one subject only. The rationale for this comparison is that high citation is a widely used indicator of research quality and hence one may expect, in general, multi-disciplinary research to be more highly cited than mono-disciplinary research. Indeed, this seems to be the case in library and information science, at least for very highly cited articles (Table 5.1).

Inter-disciplinary research can be regarded as the amalgamation of different fields into a new field, whereas multi-disciplinary research may be regarded as dealing with the same problem area from different disciplinary viewpoints. As this is a macro-level investigation, it does not investigate the extent to which the articles in the journals can be classified as inter-disciplinary research or multi-disciplinary research; it examines the subject classifications of the journals.

In order to address the third global research question, this chapter compares for different time periods, subjects and databases the citation level of multi-disciplinary journals with those of mono-disciplinary journals in order to address the following research questions on the relationship between level of citation and level of disciplinarity:

1. Are multi-disciplinary journals in science and the social sciences on average more highly cited than mono-disciplinary journals?
2. Has the citation level of multi-disciplinary journals in the social sciences relative to mono-disciplinary journals changed over time?
3. Is the citation level of the journals in a combination of two subjects related to the citation levels of the journals in the component subjects?

The citation level of a journal is defined as the level of citation of all its articles. Although both sciences and social sciences are investigated, because of practical limitations the sciences are not analysed in as much detail as the social sciences.

7.2: Methods

The research questions are investigated by comparing, for diverse subjects of the WoS and Scopus databases, the citation levels of two disjoint sets of journals that together make up all journals in a subject. For each subject, one set, called 'Mono', consists of all journals classified solely in that subject. The other set, called 'Multi', consists of the remaining journals in the subject. The data was obtained via the Internet by conducting searches on the databases; the WoS searches made extensive use of the 'Refine your results' and 'Citation Report' facilities and the Scopus searches made extensive use of the 'Refine Results' facility and the option to sort by 'Cited By'.

Question 2 is addressed by comparing findings for two different years, and the other questions are addressed by investigating a single year. For reasons discussed below the data source for question 1 is both WoS and Scopus, for question 2 solely WoS, and for question 3 solely Scopus. Note that in the text below we capitalise the subject category names in order to differentiate between the names and the subjects that they approximately represent.

The investigation of question 1 examines two subsidiary questions to provide a broader understanding of the relationship between citation and disciplinarity:

s1 Are differences in the average level of citation between multi-disciplinary and mono-disciplinary journals dependant on the subject area (e.g., are the findings the same for subjects classified as Life Sciences, Health Sciences, Physical Sciences and Social Sciences)?

s2 Are differences in the average level of citation between multi-disciplinary and mono-disciplinary journals dependant on the data source (e.g., do the findings for Scopus differ from those for WoS)?

In order to address the research questions, the simplifying assumption is made that research published in journals categorised in more than one subject by WoS or Scopus is multi-disciplinary, whereas research published in journals that are categorised in only one subject is mono-disciplinary. This is clearly an oversimplification, not least because Bradford's (1934) law of scattering implies that research is not always published in the core journals of a field, but also because some multi-disciplinary journals, such as Nature and Science, publish significant amounts of mono-disciplinary research (Ackerson & Chapman, 2003). Moreover, the subject categories of the two databases are optimised for information retrieval rather than scientometric evaluation, and the issue of identifying disciplines is complex and without an easy solution (Glänzel & Schubert, 2003). Nevertheless, the simplification used here seems like a reasonable method to differentiate between two sets of journals, one of which is likely to contain higher levels of inter-disciplinary research than the other. This assumption is supported by the Morillo, Bordons and Gómez (2001) study that found that a WoS subject with a high level of overlap with other subjects had proportionately more external citations than did a WoS subject with a low level of overlap with other subjects, which is suggestive of greater inter-disciplinarity. Moreover, numerous studies have found the WoS disciplinary categories to be a useful data source (Borgman & Rice, 1992; Qin, Lancaster, & Allen, 1997; Hinze, 1999; Morillo, Bordons, & Gómez, 2001; Rinia, Van

Leeuwen, Bruins, Van Vuren, & Van Raan, 2002; Rinia, Van Leeuwen, & Van Raan, 2002; Eto, 2003; Morillo, Bordons, & Gómez, 2003) and they are the default data source for many research evaluation exercises, such as that proposed for the U.K. (http://www.hefce.ac.uk/research/assessment/faq/, last visited September 5, 2008).

This chapter presents two citation analyses of disciplinarity. It examines disciplinarity in social science by investigating the SSCI and it examines disciplinarity in science and social science by investigating Elsevier's Scopus. The reason for using both the SSCI and Scopus for investigation of disciplinarity in social science is that this enables a comparison between the WoS and Scopus subject categories. Scopus was chosen in preference to WoS for investigating disciplinarity in science, as WoS has many more science subject categories (172 compared with 21). Collectively these investigations cover 27 Scopus subjects and 28 WoS social science subjects, thereby allowing not only comparisons between subject areas (question s1) but also comparison between databases (question s2). Using Scopus and Google Scholar, in addition to WoS, can provide a more accurate and comprehensive picture of the citation impact of influential information scientists than using WoS alone (Meho & Yang, 2007). Although Google Scholar has been found to have a higher percentage of core articles than the SSCI (Walters, 2007), Google Scholar was not investigated here because it does not have subject categories.

The SSCI investigation examines articles published in 1986. The choice of 1986 was a matter of judgement; the earlier the publication the longer the period of citation, but the later the year of publication the more likely that the findings apply to contemporary data. The investigation of Scopus examines articles published in 1995. The year 1995 was chosen because Scopus does not provide any citation data prior to 1995. For each database this study investigates the disciplinary categories that contain the most articles because findings on larger frequencies are less likely to be spurious.

This chapter uses two indicators to compare the level of citations of disciplines; the mean number of citations per article, and the Normalised Hirsch Index derived from the Hirsch Index. The Hirsch Index (Hirsch, 2005) is defined to be the largest number h of documents that are cited h or more times. This has become accepted as a reasonable indicator of the impact of a body of work (e.g., Cronin & Meho, 2006; Oppenheim, 2007). Hirsch indexes are quick to calculate for WoS and Scopus as both databases allow articles to be ranked in decreasing order of citation. However one problem with comparing Hirsch indexes is that they do not adjust for the number of documents investigated and so the Normalised Hirsch Index was used in order to overcome this problem. The Normalised Hirsch Index h_{norm} for a set of documents is defined by

$$h_{norm} = 100\, h^2 / n$$

where h and n are the h-index and number of documents of the set.

The Normalised Hirsch Index (h_{norm}) is useful for comparing the citation levels of disciplines in Scopus because Scopus does not provide data on the average number of citations per article (i.e., the more standard impact indicator). For the Web of Science it is useful to supplement the average number of citations per article, multiple indicators are preferable to single indicators as they provide more information (e.g., Martin, 1996; Van Leeuwen, Van der Wurff, & Van Raan, 2001). For all the investigations the largest possible citation window (i.e., citations to date) was used because the longer the citation window the more closely the findings are likely to approximate to eventual citations.

7.3: Results

7.3.1: Social science in the Web of Science (1986 and 1995)

This investigation examines two sets of SSCI articles. The first set consists of all articles published in 1986 in the 28 SSCI subjects in which at least 1,000 articles were published in 1986; the second set consists of all articles published in 1995 in these 28 subjects.

The 28 subjects were consolidated into 20 categories by: (a) Combining the eight different subjects for Psychology into a single category called 'Psychology (8 categories)', and (b) Combining 'Business' and 'Business Finance' into 'Business OR Business Finance'. Data on the mean number of citations, h, h_{norm} and coverage of the articles published in the 20 categories in 1986 and 1995 is presented in Table 7.1. In the table, 'Mean' is the mean number of citations per article in the category, '% of SSCI' the category's coverage as a percent of all SSCI articles for the year, and 'Other categories' is a composite category consisting of the articles published in the year that are not in any of the 28 categories.

Table 7.1: The mean number of citations, h values, h_{norm} and coverage for 20 SSCI subject categories (1986 and 1995)

	1986				1995			
Category	**Mean**	**h**	**h_{norn}**	**% of SSCI**	**Mean**	**h**	**h_{norn}**	**% of SSCI**
Anthropology	8.63	40	139	2.0	6.65	38	99	2.0
Behavioral Sciences	19.89	74	385	2.5	20.34	76	239	3.4
Business OR Business Finance	11.16	87	212	6.2	12.89	86	214	4.8
Economics	12.48	106	220	8.9	10.56	90	123	9.2
Education & Educational Research	5.84	48	73	5.5	5.72	47	66	4.6
Environmental Studies	6.25	32	101	1.8	8.36	41	115	2.0
Information Science & Library Science	3.54	32	59	3.0	3.89	37	59	3.3
International Relations	2.27	26	42	2.8	4.26	33	70	2.2
Law	6.21	57	99	5.7	5.54	45	70	4.1
Management	19.49	81	366	3.1	14.84	81	229	4.0
Neurosciences	22.93	80	512	2.2	35.60	149	520	6.0
Planning & Development	3.78	29	49	3.0	5.03	37	78	2.5
Political Science	2.94	41	45	6.5	3.89	43	51	5.1
Psychiatry	24.75	129	406	7.1	23.49	128	279	8.2
Psychology (8 categories) *	19.70	166	227	21.1	17.11	141	127	21.9
Public, Environmental & Occupational Health	18.32	77	304	3.4	16.92	85	218	4.6
Rehabilitation	9.69	44	170	2.0	9.72	44	118	2.3
Social Issues	2.40	25	45	2.4	4.05	26	62	1.5
Social Sciences, Interdisciplinary	3.82	37	62	3.8	5.77	40	87	2.6
Sociology	10.97	62	193	3.4	8.50	54	114	3.6
Other categories	10.90	125	121	22.4	12.17	141	127	21.8
Mean	**10.76**	**67**	**182**	**5.7**	**11.20**	**70**	**146**	**5.7**

* The 8 categories are 'Psychology', 'Psychology, Applied', 'Psychology, Clinical', 'Psychology, Developmental', 'Psychology, Educational', 'Psychology, Experimental', 'Psychology, Multidisciplinary' and 'Psychology, Social'**.**

In Table 7.1, the mean number of citations per article for 1986 ranges from 2.27 for International Relations to 24.75 for Psychiatry and for 1995 ranges from 3.89 for Information Science & Library Science to 35.60 for Neurosciences 57,689 SSCI articles were published in 1986 and 71,841 SSCI articles were published in 1995. For 1986 the Pearson correlation between the mean number of citations and h-index was .75 and the correlation between the mean number of citations and h_{norn} was .96. This very strong correlation indicates that the Normalised Hirsch Index is a strong indicator of average number of citations and so it is reasonable to use h_{norm} to substitute for average citations when data on the mean number of citations is unavailable.

Table 7.2 compares indicators for Mono and Multi journals. This data is used to examine questions 1 and 2. 'Mono as % of all' denotes the percentage of Mono articles, 'Mean Mono' and 'Mean Multi' denote the mean numbers of citations for the articles in Mono and Multi journals, 'h Mono' and 'h Multi' denote the h-values for the articles in Mono and Multi, and 'h_{norm} Mono' and 'h_{norm} Multi' are the values of h_{norm} for the articles in Mono and Multi.

Table 7.2: Comparison of mean number of citations per article and h_{norm} for mono- and multi-disciplinarity (SSCI 1986)

Category	Mono as % of all	Mean Mono	Mean Multi	h Mono	h Multi	h_{norm} Mono	h_{norm} Multi
Anthropology	50.1	5.86	11.41	25	37	109	239
Business OR Business Finance	46.2	7.99	13.89	53	73	170	278
Economics	49.2	13.19	11.80	83	72	274	200
Education & Educational Research	67.3	5.20	7.17	40	34	75	112
Environmental Studies	14.2	2.69	6.84	10	31	69	110
Information Science & Library Science	55.6	2.47	4.88	15	29	23	108
International Relations	32.6	3.03	1.90	17	20	55	37
Law	58.0	8.14	3.54	54	26	153	49
Planning & Development	15.0	4.42	3.67	13	28	66	54
Political Science	38.5	4.40	2.02	36	28	89	34
Psychiatry	46.4	28.68	21.35	111	90	647	368
Psychology (8 categories)	61.5	21.41	16.98	151	100	305	214
Public, Environmental & Occupational Health	41.1	20.56	16.76	60	59	449	303
Rehabilitation	29.9	5.50	11.47	20	44	117	242
Social Issues	21.2	3.70	2.05	16	21	87	40
Social Sciences, Interdisciplinary	21.5	2.82	4.09	15	37	47	79
Sociology	43.2	11.90	10.26	48	45	268	179
Neurosciences	6.8	41.98	21.54	28	73	922	457
Management	2.6	19.81	19.48	16	80	545	367
Behavioral Sciences	0	NA	19.89	NA	74	NA	385

In Table 7.2 the percentage of articles in Mono journals ranges from 0% for Behavioural Sciences to 67.3% for Education & Educational Research. For the 17 categories in which the percentages of Mono and Multi are both more than 10%, the mean number of citations averages 8.94 for Mono and 8.83 for Multi. This evidence is not supportive of the hypothesis that, in general, articles in multi-disciplinary journals are cited more often on average than articles in mono-disciplinary journals. For the 10 categories where the mean number of citations for Mean Mono exceeds that of Mean Multi the ratio of Mean Mono to Mean Multi averages 1.52; for the other 7 categories the ratio of Mean Multi to Mean Mono averages 1.88. For the 17 categories, the Normalised Hirsch Index averages 176.6 for Mono and 155.6 for Multi. The findings presented in Table 7.2 are not supportive of the hypothesis that high citation is more prevalent amongst articles in multi-disciplinary journals than in articles in mono-disciplinary journals.

A question that can be asked regarding Table 7.2 is: Do a small number of journals in a subject account for much of the differences between Mono and Multi? Answering this question for all the subjects would require extensive further research. Here this question is addressed for the two subjects which have the highest and lowest ratios of Mean Mono to Mean Multi. The method used is to identify for each subject journals that contain at least one of the six most cited articles and examine the impact of these journals on Mean Multi and Mean Mono. In Table 7.2, the highest ratio of Mean Mono to Mean Multi is 2.30 for Law. For Mono the six most cited articles are in five journals that together form 9.5% of Mono articles and average 31.16 citations per article; for Multi the six most cited articles are in five journals that together contain 12.5% of the Multi articles and on average have 11.74 citations per article. Excluding for both Mono and Multi the journals with the six most highly cited articles, the ratio of Mean Mono to Mean Multi is 2.30 (Mono averages 5.66 and Multi 2.46). In Table 7.2, the lowest ratio of Mean Mono to Mean Multi is .39 for Environmental Studies. The six most cited Mono articles are in four journals that together form 52.1% of Mono articles and average 4.19 citations per article. The six most cited Multi articles are in five journals that together form 25.3% of Multi articles and average 10.23 citations per article. Excluding for both Mono and Multi the journals with the six most highly cited articles, the ratio of Mean Mono to Mean Multi is .19 (Mono averages 1.13 and Multi 5.88). Thus for Law excluding the journals containing the six most highly cited articles resulted in no change to the high ratio of Mean Mono to Mean Multi whereas for Environmental Studies the low ratio of Mean Mono to Mean Multi became lower when the journals were excluded.

Table 7.3 presents similar data to Table 7.2 for articles published in 1995. This is used to examine questions 1, 2 and s2. The percentage of articles solely in a category ranges from 0% for Behavioural Sciences to 62.6% for Education & Educational Research. For the 17 categories for which averages were obtained for the 1986 data (all categories except for Neurosciences, Management and Behavioral Sciences) the mean number of citations averages 8.27 for Mono and 9.09 for Multi. This evidence is again not supportive of the hypothesis that, in general, articles in journals in multiple SSCI categories are cited more often on average than articles in a single category. For the 9 categories where the mean number of citations for Mono exceeds that for Multi (the categories in Table 7.2 apart from Psychology for which Mean Mono exceeds Mean Multi) the ratio of Mono to Multi averages 1.38; for the other 8 categories the ratio of Multi to Mono averages 2.00. For the 17 categories, the Normalised Hirsch Index averages 153.3 for Mono and 130.4 for Multi; this latter data is not consistent with high citation being more prevalent amongst articles in multiple disciplines than in articles in a single discipline. Interestingly, for both Tables 7.2 and 7.3, for each category in if Mean Mono exceeds Mean Multi then h_{norm} Mono exceeds h_{norm} Multi and if Mean Multi is less than Mean Multi then h_{norm} Mono is less than h_{norm} Multi: This indicates that h_{norm} can be used as an alternative indicator to the average number of citations. Using the same notation as Table 7.2, 'Mono as % of all' denotes the percentage of Mono articles, 'Mean Mono' and 'Mean Multi' denote the mean numbers of citations for the articles in Mono and Multi journals, 'h Mono' and 'h Multi' denote the h-values for the articles in Mono and Multi, and 'h_{norm} Mono' and 'h_{norm} Multi' are the values of h_{norm} for the articles in Mono and Multi.

Table 7.3: Comparison of mean number of citations per article and h_{norm} for mono- and multi-disciplinarity (SSCI 1995)

Category	Mono as % of all	Mean Mono	Mean Multi	h Mono	h Multi	h_{norm} Mono	h_{norm} Multi
Anthropology	38.5	3.66	8.22	18	36	58	144
Business OR Business Finance	42.8	9.62	14.47	54	74	197	276
Economics	46.1	10.40	10.08	74	69	181	134
Education & Educational Research	62.6	4.58	7.24	37	36	66	104
Environmental Studies	11.9	3.69	8.74	12	40	83	124
Information Science & Library Science	53.0	1.65	6.15	16	36	21	118
International Relations	31.8	6.08	3.30	26	21	136	42
Law	54.5	5.75	5.13	39	32	96	77
Planning & Development	11.3	5.26	4.85	14	35	99	78
Political Science	46.0	5.16	2.64	39	25	91	32
Psychiatry	36.3	25.80	21.35	105	99	517	262
Psychology (8 categories)	58.8	16.00	17.54	121	103	158	164
Public, Environmental & Occupational Health	37.2	17.57	15.87	69	66	385	209
Rehabilitation	34.4	6.63	10.98	25	42	111	164
Social Issues	18.0	6.85	3.34	19	21	183	49
Social Sciences, Interdisciplinary	29.0	3.59	6.45	19	38	67	110
Sociology	43.7	8.30	8.23	42	43	158	129
Neurosciences	21.3	72.29	24.81	132	101	1915	303
Management	17.1	9.37	15.23	32	76	208	243
Behavioral Sciences	0	NA	19.87	NA	75	NA	233

Statistical tests were conducted to identify the subjects in Table 7.3 for which there is a correlation at the journal level between citation level and disciplinary (whether Mono or Multi). For every category in Table 7.3, apart from

Behavioral Sciences, for both Mono and Multi, journals satisfying the following criteria were examined: (a) At least 10 articles were published in the journal in 1995, and (b) The journal was in the list of the 30 journals with most articles published in 1995. In the case of Neurosciences, 73% of the articles in Mono journals were published in a single journal, and because of this very high percentage it was decided not to conduct further statistical tests on Neurosciences. For the remaining subjects, the Kolmogorov-Smirnov Normality test was applied for both mean citation and h_{norm}, to the sets of journals meeting criteria (a) and (b). As the condition for applying the t-test ($p > .05$ for both Mono and Multi) held only for three cases (mean citation Anthropology, mean citation Environmental Studies and h_{norm} Environmental Studies), the Mann-Whitney U test was applied throughout. The findings are presented in Table 7.4, where "Mono' and 'Multi' denote the number of Mono and Multi journals used in the tests.

Table 7.4: Ratios of Multi to Mono for mean citation and h_{norm} together with p values (Mann-Whitney U test) for the significance of the difference between them (SSCI 1995)

	Mean Citation		h_{norm}		Number of journals	
Subject category	**Multi/ Mono**	**p**	**Multi/ Mono**	**p**	**Mono**	**Multi**
Information Science & Library Science	3.73	.01	5.71	.02	30	28
Environmental Studies	2.37	.01	1.50	.01	6	30
Anthropology	2.25	.01	2.51	.01	24	28
Social Sciences, Interdisciplinary	1.80	.04	1.63	.09	18	30
Rehabilitation	1.66	.06	1.48	.03	16	30
Management	1.63	.10	1.17	.13	15	30
Education & Educational Research	1.58	.04	1.58	.02	30	30
Business *OR* Business Finance	1.50	.94	1.40	.95	30	30
Psychology (8 categories)	1.10	.50	1.04	.48	30	30
Sociology	.99	.97	.82	.84	30	30
Economics	.97	.62	.74	.76	30	30
Planning & Development	.92	.30	.79	.37	8	29
Public, Environmental & Occupational Health	.90	.82	.54	.87	25	30
Law	.89	.15	.81	.23	30	30
Psychiatry	.83	.08	.51	.04	30	30
International Relations	.54	.54	.30	.70	16	30
Political Science	.51	.49	.35	.54	30	30
Social Issues	.49	.75	.27	.94	7	29

Table 7.4 indicates an asymmetry, in that correlation between citation level and disciplinarity at the journal level is strongly associated with high ratios of Multi to Mono, but not strongly associated with high ratios of Mono to Multi. The lack of statistical significance for Management, Rehabilitation (mean citation), and Social Sciences Interdisciplinary (h_{norm}) might be due to the small number of journals examined. A total of 7 of the 18 categories had at least one p value less than .05. Of the five categories with p-values for mean citation less than .05, four had the highest ratios of Mean Multi to Mean Mono in Table 7.4 (Anthropology; Environmental Studies; Information Science & Library Science;

Social Sciences, Interdisciplinary) and the other (Education & Educational Research) had a ratio of Mean Multi to Mean Mono of 1.58. Of the six categories with p-values for h_{norm} less than .05, two had the highest ratios of h_{norm} Multi to h_{norm} Mono (Anthropology; Information Science & Library Science), three had the fourth, fifth and sixth highest ratios of h_{norm} Multi to h_{norm} Mono (Education & Educational Research; Environmental Studies; Rehabilitation) and one (Social Issues) had a ratio of h_{norm} Multi to h_{norm} Mono of .51.

The discussion examines the findings of this section in relationship to the research questions. The discussion also compares the findings for 1986 with those for 1995 (Table 7.8) and the findings for WoS with those for Scopus (Table 7.9).

7.3.2: Science and social science in Scopus (1995)

Table 7.5 presents data on all the 27 subjects and 4 subject areas of Scopus for all articles published in 1995. This data is used to examine questions 1, s1 and s2. In Table 7.5 the terminology is the same as before except that 'n Mono' and 'n Multi' denote the number of articles solely in the subject or articles in the subject and at least one other subject, respectively. The mean number of citations is not used as an indicator in Table 7.5, as Scopus does not provide this data. Note that the subject descriptors are partially misleading: for example mathematics is not a physical science and the arts and humanities are not social sciences. Using the same notation as Tables 7.2 and 7.3, 'Mono as % of all' denotes the percentage of Mono articles, 'Mean Mono' and 'Mean Multi' denote the mean numbers of citations for the articles in Mono and Multi journals, 'h Mono' and 'h Multi' denote the h-values for the articles in Mono and Multi, and 'h_{norm} Mono' and 'h_{norm} Multi' are the values of h_{norm} for the articles in Mono and Multi.

Table 7.5: Comparison of the h-indexes and Normalised Hirsch Indexes for mono- and multi-disciplinarity (Scopus 1995)

Subject	Mono articles	Mono as % of all	h Mono	h Multi	h_{norm} Mono	h_{norm} Multi
Agricultural and Biological Sciences	18,464	41.6	118	149	75	86
Biochemistry, Genetics and Molecular Biology	52,834	45.4	306	291	177	133
Immunology and Microbiology	12,233	34.3	197	162	317	112
Neuroscience	9,811	35.5	188	155	360	135
Pharmacology, Toxicology and Pharmaceutics	12,278	33.1	112	115	102	53
Life Sciences Mean	**21,124**	**40.5**	**184.2**	**174.4**	**206.4**	**103.8**
Dentistry	3,219	73.1	68	40	144	135
Health Professions	1,233	10.4	96	108	747	109
Medicine	168,661	68.4	312	276	58	98
Nursing	3,307	48.2	98	66	290	123
Veterinary	2,674	49.6	45	57	76	119
Health Sciences Mean	**35,819**	**65.1**	**123.8**	**109.4**	**263.0**	**116.8**
Chemical Engineering	4,226	22.7	42	93	42	60
Chemistry	13,837	35.8	175	122	221	60
Computer Science	7,718	27.8	65	118	55	70
Earth and Planetary Sciences	5,993	28.5	110	121	202	97
Energy	2,904	26.3	23	65	18	52
Engineering	68,031	54.3	109	138	18	33
Environmental Science	5,603	19.4	101	135	182	79
Materials Science	22,413	49.2	121	98	65	42
Mathematics	3,424	36.3	40	78	47	101
Physics and Astronomy	17,723	32.8	187	131	197	47
Physical Sciences Mean	**15,187**	**40.0**	**97.3**	**109.9**	**104.7**	**64.1**
Arts and Humanities	10,157	92.9	13	33	2	140
Business, Management and Accounting	675	15.4	17	37	43	37
Decision Sciences	167	7.1	17	47	173	101
Economics, Econometrics and Finance	505	28.3	21	56	87	245
Psychology	3,302	30.2	101	128	309	215
Social Sciences	2,742	23.8	68	88	169	88
Social Sciences Mean	**2,925**	**41.9**	**39.5**	**64.8**	**130.4**	**137.6**

For the subjects in Table 7.5 the percentage of articles solely in a subject varies from 7.1% for Decision Sciences to 92.9% for Arts and Humanities. For all categories, apart from Decision Sciences and Arts and Humanities (both in the Social Sciences category), the percentages of articles in Mono and Multi both exceed 10%. For Life Sciences, Health Sciences and Physical Sciences, the Normalised Hirsch Index for Mono was substantially larger than that for Multi with the average value of 191.4 for Mono and 94.9 for Multi. In the case of the subject area of Social Sciences the Normalised Hirsch Index for Mono is 94.8% of Multi. If the categories of Decision Sciences and Arts and Humanities are excluded from Social Sciences, Mono is 103.9% of Multi (Mono averages 151.9 and Multi 146.2).

A set of statistical tests was conducted to verify the conclusions drawn from Table 7.5. Only social sciences were examined due to the prohibitively long time needed to collect the necessary data. For every subject in Social Sciences in Table 7.5, for both Mono and Multi, journals that satisfy the following criteria were examined: (a) At least 10 articles were published in the journal in 1995, and (b) The journal was one of the 30 journals with most articles published in 1995. In the case of Decision Sciences, 66% of the articles in Mono journals were published in a single journal, and because of this very high percentage it was decided not to conduct further statistical tests on Decision Sciences. For the remaining subjects, the Kolmogorov-Smirnov normality test was again applied for h_{norm} to the sets of journals meeting criteria (a) and (b). As the condition for applying the t-test ($p > .05$ for both Mono and Multi) held only for one subject (Psychology), the Mann-Whitney U test was applied throughout. The findings are presented in Table 7.6.

Table 7.6: Statistical significance using the Mann-Whitney U test of the differences between Mono and Multi h_{norm} values (Scopus 1995)

Subject	Signifi-cance p	Mono journals	Multi journals
Arts and Humanities	.00	30	30
Business, Management and Accounting	.92	21	30
Economics, Econometrics and Finance	.04	14	30
Psychology	.13	30	30
Social Sciences	.00	30	30

In Table 7.6, two of the three subjects with p values less than .05 were the only social science subjects in Table 7.5 for which h_{norm} Multi exceeds h_{norm} Mono (Arts and Humanities; Economics, Econometrics and Finance).

This section next investigates the level of citation of subjects that overlap. Table 7.7 presents data on the 9 subject combinations that contain over 10,000 articles (prior to excluding articles in more than two subjects). This data is used to examine question 3. The hypothesis underlying this investigation is that the Normalised Hirsch Index for articles in both subject A and subject B is very roughly the mean of the Normalised Hirsch Indexes for subject A and subject B. In the table, 'n Comb', 'h Comb' and 'h_{norm} Comb' denote the number of articles, h-index and Normalised Hirsch Index of the subject combination, and 'Mean h_{norm} of the components' denotes the average of the h_{norm} of the two component subjects of the subject combination. Note that 'Articles' is often lower than 10,000 as it does not include any article in more than two subjects.

Table 7.7: Comparison of the Normalised Hirsch Indexes for subject combinations with those for their component subjects (Scopus 1995)

Combination	Articles	h Comb	h_{norm} Comb	Mean h_{norm} of the components
Biochemistry, Genetics and Molecular Biology *AND* Agricultural and Biological Sciences	4,027	108	290	126
Medicine *AND* Biochemistry, Genetics and Molecular Biology	22,169	249	280	118
Medicine *AND* Immunology and Microbiology	10,525	147	205	188
Medicine *AND* Neuroscience	6,449	113	198	209
Medicine *AND* Health Professions	5,795	93	149	403
Biochemistry, Genetics and Molecular Biology *AND* Pharmacology, Toxicology and Pharmaceutics	2,607	54	112	140
Engineering *AND* Computer Science	11,943	94	74	36
Physics and Astronomy *AND* Materials Science	7,189	69	66	131
Engineering *AND* Physics and Astronomy	16,694	76	35	107

In Table 7.7, the average of the Normalised Hirsch Index for the subject combination is 156.6 and the average of the mean of component subjects is 162.0. For the four cases in which the Normalised Hirsch Index of the subject combination exceeds that of the mean of the component subjects the ratios of 'h_{norm} Comb' to 'mean h_{norm} of the components' average 1.96. For the five cases in which the Normalised Hirsch Index of the mean of the component subjects exceeds that of the subject combination the ratios of 'mean h_{norm} of the components' to 'h_{norm} Comb' averages 2.01.

7.4: Discussion

The Mann-Whitney U and t-tests at the journal level identified that statistically significant results were more prevalent when Multi exceeded Mono than when Mono exceeded Multi. The link between statistically significant results and high ratios of Multi to Multi was particularly strong in that: (a) For mean citation the four categories with the highest ratio of Multi to Mono all had statistical significance and yet only one of the remaining 14 categories had statistical significance, and (b) For h_{norm} five of the 6 categories with the highest ratio of Multi to Mono had statistical significance and yet only one of the remaining 12 categories had statistical significance. We conjecture that this difference between Mono and Multi probably also applies to other subjects in WoS and to other years.

The findings from Tables 7.2 and 7.3 are compared in Table 7.8. This indicates that for both 1986 and 1995 the averages of the mean number of citations for Multi is close to that of Mono (values of Multi/Mono are .99 and 1.10). For both 1986 and 1995 the average Normalised Hirsch Index for Multi is close to that of Mono (values of Multi/Mono are .88 and .85). These findings answer questions 1 and 2, finding at the subject level: (a) Multi-disciplinary journals were not on average significantly more highly cited than mono-disciplinary journals, and (b) For the 16 categories apart from Psychology, whenever Mean Multi exceeds Mean Multi in 1986 it also exceeds it in 1995 and whenever Mean Multi is less than Mean Multi in 1986 it is also is less in 1995. There were considerable variations between subjects, for both 1986 and 1995 for both Political Science and Social Issues the ratio of Mean Mono to Mean Multi is less than .52, whereas for both Anthropology and Information Science & Library Science the ratio exceeds 2.18.

For both periods the ratio of Multi to Mono when Multi is greater is considerably larger than the ratio of Mono to Multi when Mono is greater. Table 7.8 also indicates that the dissymmetry between these ratios was noticeably larger for 1995 than for 1986 (2.00/1.38 compared with 1.88/1.52). These findings indicate that the asymmetry between Multi and Mono has increased over time.

Table 7.8: Comparison between the findings for SSCI articles published in 1986 and 1995

Category	1986	1995
Average of mean citations for Mono	8.94	8.27
Average of mean citations for Multi	8.83	9.09
Ratio of Mono to Multi when Mono is greater	1.52	1.38
Ratio of Multi to Mono when Multi is greater	1.88	2.00
Average Normalised Hirsch Index for Mono	176.6	153.3
Average Normalised Hirsch Index for Multi	155.6	130.4

In Table 7.8 the citation window for articles published in 1986 is about nine years longer than those for 1995, as in both cases the citation window is to date. The longer citation window could account for 1986 having higher values than 1995 for the average number of citations and Normalised Hirsch Indexes. Although averages provide a broad indication of how Mono and Multi compare, they seem likely to be skewed by subjects that have high citation rates. A comparison of the ratio of Mono to Multi for individual categories yields an interesting finding: The same 7 subject categories show higher citation rates for Multi than for Mono in both 1986 and 1995, while the opposite hold for 9 categories. This finding needs to be treated with caution, as the length of citation window for 1986 differs from that for 1995.

In Table 7.4 whilst journal level correlations between citation level and disciplinarity (Mono or Multi) are strongly associated with high ratios of Multi to Mono, these journal level correlations are not strongly associated with high ratios of Mono to Multi. In order to understand this asymmetry, revised WoS ratios of Mean Multi to Mean Mono were calculated for the three cases for which the ratio of Mean Multi to Mean Mono was highest (Information Science & Library Science, Environmental Studies, Anthropology) and the three subjects for which the ratio was lowest (International Relations, Political Science, Social Issues). The revised technique used was to exclude journals with the highest number of citations, with the number of journals excluded being proportional to the number of journals previously examined (e.g., 4 excluded from Multi Environmental Studies, 3 excluded from Multi Anthropology, 2 excluded from Mono International Relations and 1 excluded from Mono Social Issues).

The revised ratios of Mean Multi to Mean Mono were 2.44 for Information Science & Library Science, 2.16 for Environmental Studies, 1.87 for Anthropology, .66 for International Relations, .68 for Political Science and .93 for Social Issues. Where Mean Multi exceeded Mean Mono, the average ratio of Mean Multi to Mean Mono in the revised ratios reduced somewhat to 2.16 compared with the previous value of 2.78; where Mean Mono exceeded Mean Multi, the average ratio of Mean Multi to Mean Mono in the revised ratios increased somewhat to .75 compared with the previous value of .51.

In order to compare the findings of Scopus with WoS, some of the results in Tables 7.5 and 7.8 are presented in Table 7.9. This answers question 1: For the data investigated in Life Sciences, Health Sciences and Physical Sciences mono-disciplinary journals were on average substantially more highly cited than multi-disciplinary journals (on average h_{norm} for Mono was 201% of h_{norm} for Multi). The table also answers question s1: For the data investigated the findings on the relationship between citation level and disciplinarity for Social Sciences differed substantially from those for the other Scopus Subject areas.

Table 7.9: Comparison of the average Normalised Hirsch Index between Scopus subject areas and the SSCI (articles published in 1995)

Data source	Mono	Multi
Life Sciences (Scopus)	206.4	103.8
Health Sciences (Scopus)	263.0	116.8
Physical Sciences (Scopus)	104.7	64.1
Social Sciences (Scopus)	130.4	137.6
Social Sciences excluding subjects with Mono or Multi less than 10% (Scopus)	151.9	146.2
17 SSCI categories	153.3	130.4

Table 7.9 indicates that for the Scopus subject area of Social Sciences the average Normalised Hirsch Index for Mono is similar to that for Multi (94.8% for all subjects or 103.9% when subjects for which the percentage of Mono or Multi articles is less than 10% are excluded); these percentages are similar to those for the 17 SSCI categories (117.6%). This finding answers question s2 for the data investigated: For Social Science articles published in 1995 the WoS findings are similar to those for Scopus. However, there are large differences between subjects and the comparison between the SSC and Scopus does not take into account differences in the ways the databases classify subjects.

One problem with making deductions based on these averages is that the subjects of the two databases differ substantially, for instance, there is no WoS equivalent to Scopus' Decision Sciences. Whilst it might seem attractive to confine comparisons between WoS and Scopus to subjects defined in both databases, there are still problems with differences in coverage: For 1995 the WoS category of Psychology has more than four times as many articles as Scopus' Psychology (15,715 compared with 3,302) and WoS' Economics is about three times as large as Scopus' Economics, Econometrics and Finance (5,113 compared with 1,711). Differences in coverage could account for Mono being higher than Multi for WoS' Economics and Multi being higher than Mono for Scopus' Economics, Econometrics and Finance.

Question 3 asks whether the citation level of the journals in a combination of two subjects is related to the citation levels of the journals in the component subjects. Taking all 9 subjects together, the data in Table 7.7 indicates that the average of the Normalised Hirsch Index for the subject combination is 156.6 and the average of the mean of component subjects is 162.0. However there was considerable variation between subjects: For the 4 cases where the Normalised Hirsch Index of the subject combination exceeds the mean of the component subjects the ratios of 'h_{norm} Comb' to 'mean h_{norm} of the components' average 1.96; for the other 5 cases the ratios of 'mean h_{norm} of the components" to 'h_{norm} Comb' averages 2.01. Thus although for the 9 subjects together the average h_{norm} is within 4% of the mean h_{norm} of the component subjects, the h_{norm} for the individual subjects differ on average by roughly a factor of two from the mean h_{norm} of the component subjects.

7.5: Concluding Remarks

The results suggest that research in multi-disciplinary journals is not more highly cited than research in mono-disciplinary journals. This does not support the assertion that, in general, multi-disciplinary articles are more highly cited than mono-disciplinary articles. On the contrary, the major difference between mono-disciplinarity and multi-disciplinarity indicates higher citation of mono-disciplinary articles: On average for Life Sciences, Health Sciences and Physical Sciences the level of citation for mono-disciplinary was more than double that for multi-disciplinary. These results contrast with the finding in Chapter 5 that all except 2 of the 77 most highly cited articles in IS&LS articles are multi-disciplinary. Perhaps IS&LS is an exception in producing particularly high quality multi-disciplinary research, but it is also possible that multi-disciplinarity is an advantage for the highest quality research, but not for average research.

This study found statistically significant correlations at the journal level between citation level and disciplinarity in WoS 1995 when high ratios of Mono to Multi were present. When approximately one seventh of the journals with the most citations were excluded, for the three cases with the highest ratio of Mean Multi to Mean Mono declined by 22% and for the three cases with the lowest ratio it declined by 32%.As described in Chapter 9, this study has limitations; in particular, higher citation does not necessarily indicate that mono-disciplinary articles are, in general, of a higher quality than multi-disciplinary articles. Nevertheless, a clear policy implication of the findings is that the promotion of Mode 2 research is unlikely to be reflected in higher citation rates. Moreover, if

Mode 2 research is accepted as valuable, despite its apparently lower citation rate, then Mode 2 researchers – and all inter-disciplinary researchers – should not be penalised for having lower citation rates than other researchers.

The third global research question asks whether, in general, articles in multi-disciplinary journals are more highly cited than articles in mono-disciplinary journals. In Chapter 7 the main findings are: (a) For social science subject categories in both the WoS and Scopus, the average citation levels of articles in Multi and Mono are very similar, and (b) For Scopus subject categories within Life Sciences, Health Sciences, and Physical Sciences, the average citation level of Multi articles is roughly half that of Mono articles.

Finding (a) (dissimilar to finding (a) in the Concluding Remarks of Chapter 5) indicates that in social science multi-disciplinary research is not necessarily more highly cited than mono-disciplinary research; finding (b) indicates that for Scopus in science, in general multi-disciplinary research is substantially less highly cited than mono-disciplinary research. These findings indicate that there is not a clear-cut association between multi-disciplinary research and research quality, an assumption underlying the current emphasis on promoting multi-disciplinary research.

Chapter 8: Citation patterns in science and social science

Chapter 8 addresses the fourth global research question: In general, do very highly cited articles have different patterns of annual citation from less highly cited articles? It consists of five sections: Introduction, Methods of selecting cases and data processing, Results, Discussion and Concluding remarks. The limitations shared with other chapters are presented in Chapter 9 and the related further research is presented in Chapter 10. Chapter 8 presents the research in Levitt and Thelwall (in press 2008b). It relates to Chapter 5, in that the latter chapter examines the patterns of annual citation of the most highly cited 77 IS&LS articles.

8.1: Introduction

Prediction of citation ranking is implicit in citation analysis's evaluations of individuals, organisations and countries, for example when evaluating promotion, tenure, and the allocation of research funding. Hence it is important to examine whether the standard indicator (the sum of early citations) is always the most accurate method for predicting citation ranking. Additionally, if citation is an indication of research influence, then citation patterns may indicate how that influence has changed with time. This study addresses these issues using six conjectures to examine sets of highly cited articles in diverse WoS subjects that contain articles experiencing late citation.

It also seeks to identify an indicator that, for the cases examined, is a more accurate predictor than the standard indicator of total citation ranking and total citation. Here, 'early citation ranking' is used to denote the ranking of a paper based on its citations during the first few years after publication and 'total citation ranking' to denote the ranking of a paper based on its total citations. Total citation ranking can change with time in that it can be affected by future citation; early citation ranking is not unique as it depends upon the time interval classified as 'early' and also on the ranking indicator.

This chapter addresses the fourth global research question by examining six conjectures that are related to the literature.

That the pattern of citation of highly cited articles varies from article to article was identified more than 20 years ago. Garfield (1985a, 1985b) presents graphs that compare patterns of citation of highly cited papers. Garfield's graphs contain at least two different citation patterns: (a) Rising to a peak and then a steady decline, and (b) Continuing increase in citation level. This study investigates whether there are, not only two patterns, but also considerable diversity in the patterns of citation of very highly cited articles (Conjecture 1).

Recapping on Section 2.4, several studies have quantified the late citation of highly cited articles. Aversa (1985) examined the patterns of citation of 400 papers published in 1972 that were cited 30 or more times between 1972 and 1977 and found that an early rise in citation is associated with a more rapid decline in citation and a lower citation total, whereas a delayed rise in citation is associated with a less rapid decline in citation and a higher citation total. McCain and Turner (1989) compared four slowly aging and seven rapidly aging highly cited articles in Molecular Genetics. Regarding the findings, McCain (2007) wrote, "Late-peaking, slowly aging papers were likely to be cited for important widely useful methodologies or fruitful, broadly relevant theoretical insights while the early-peaking, quickly-aging papers were being cited primarily for their explicit research results." Cano and Lind (1991) compared the annual citation patterns of ten highly cited papers with ten papers with low citation in medicine and biochemistry. They found two types of citation patterns, A and B. The number of citations in the first six years, as a percentage of total citations, was generally 75% for Type A articles and 33% for Type B articles. Six of the highly cited, but none of those with a low citation level were of Type B. All six were classified as "sharing a methodological nature." Aksnes (2003) examined the patterns of annual citation of 137 highly cited papers in Norwegian science published between 1981 and 1989 and found that 33% of the papers in Physical, Chemical and Earth Sciences had the citation pattern of 'Early rise & rapid decline', whereas none of the papers in Biology and Environmental Sciences had this citation pattern. One pattern of citation referred to by both Merton (1968) and Garfield (1975) is the 'obliteration phenomenon'. For example Garfield (1993) referred to 'obliteration by incorporation', in which "discoveries or ideas become so fully incorporated into canonical knowledge that their source is no longer cited or even alluded to." In contrast, Glänzel, Schlemmer and Thijs (2003) and Van Raan (2004) established the frequencies of types of late citation that they termed respectively 'delayed recognition' and 'sleeping beauties'. They found that 0.3% of those papers that were cited more frequently than 15 times in total had not been cited between 1980 and 1984; Van Raan found that only 41 of the articles from the ISI Citation Indexes published in 1988 received at most ten citations during the first ten years after publication and subsequently between 21 and 30 citations in the next four years. Other studies of delayed recognition include those of Garfield (1980) and Glänzel and Garfield (2005) and other studies of sleeping beauties include those of Van Dalen and Henkens (2005) and Burrell (2005).

To add to the general knowledge of citation patterns developed by the above mentioned researchers, the present study takes a multiple subject comparison approach and investigates whether for a sufficiently high citation level:

- There is, for most subjects, considerable variation between articles in the percentage of citations that occur in the first few (e.g., 6) years after publication (Conjecture 2). This provides more fine-grained information on the variation in early citation than do Aversa's (1985) and Aksnes' (2003) categorisations of early citation.
- For some subjects the correlation between the percentage of early citations and total citation ranking is strong, whereas for other subjects the correlation is weak (Conjecture 3). This follows up, at the subject level, a topic investigated by Aversa (1985) who found that a delayed rise in citation is associated with a higher citation total.
- For all subjects the mean citation pattern of the highly cited articles differs substantially from the mean citation pattern of the remaining articles in the subject (Conjecture 4). This builds on Cano and Lind (1991) who

compared the citation patterns of ten highly cited papers with those of ten papers with a low citation level.

- Late citation peaks are present amongst a wide diversity of subjects or subject combinations (Conjecture 5). This builds on the finding in Chapter 5 that late citation peaks are common amongst the 77 most highly cited articles in IS&LS.

The conjectures focus on the concepts of early citation, patterns of citation and late citation that have been previously investigated. One feature of this current investigation is that it compares highly cited articles in different subjects. Aksnes (2003) and Adams (2005) share this feature, but they confine their investigations to articles by authors from a single country. Because this current investigation compares different subjects and does not limit its articles to authors from specific countries, its findings have greater generality than previous investigations that are confined to a single subject or to authors from a single country. Although the possible relationship between type of paper and late citation suggested by McCain (2007) is interesting, that investigation is outside the scope of this thesis.

This chapter also relates to the research of Adams (2005) on the correlation between early and late citation ranking. For different subject areas within the U.K., Adams examined the Pearson correlation between the number of citations in the year after publication with the total number of citations in the subsequent ten years. Adams found that for both life and physical sciences "the most highly ranked papers initially will remain amongst the higher ranked papers on average", but in the physical sciences "some papers with high total impact may initially gain less recognition." The present chapter investigates whether, for a sufficiently high citation level, prediction of total citation ranking and total citations is much more accurate in some subjects than in other subjects (Conjecture 6).

8.2: Methods of selecting cases and data processing

This chapter examines six cases, chosen as particularly likely to indicate differing citation patterns in science and social science.

Three cases investigate articles classified as 'Physics, Atomic, Molecular & Chemical', 'Chemistry, Multidisciplinary' and 'Physiology' from the SCI. These subjects were chosen for having at least one article represented amongst the 54 SCI articles that were published in 1970 and that have been cited at least 1,000 times (here referred to as the 'SCI set', data presented in Table 8.1). The first two subjects were chosen on the basis of late citation peaks and high percentages of citations during 1995 to 2006, whereas the last subject was chosen on the basis of an early citation peak and a low percentage of citations during 1995 to 2006. Of the SCI articles published in 1969–71, 7,186 were categorised as 'Physics, Atomic, Molecular & Chemical', 42,576 as 'Chemistry, Multidisciplinary' and 11,901 as 'Physiology'.

The other three cases investigate articles classified as 'Economics', 'Statistics & Probability' and 'Psychology, Multidisciplinary' from the SSCI. These subjects were chosen on the basis of at least one article represented amongst the 33 SCI articles that were published in 1970 that have been cited at least 500 times (here referred to as the 'SSCI set', data presented in Table 8.1). The first two subjects were chosen on the basis of late citation peaks and high percentages of citations during 1995 to 2006, whereas the last subject was chosen on the basis of an early citation peak and a low percentage of citations during 1995 to 2006. Of the SSCI articles published in 1969–71, 8,592 were categorised as 'Economics', 2,419 as 'Statistics & Probability' and 4,741 as 'Psychology, Multidisciplinary'.

For both the SCI and SSCI cases, comparison between subjects needs to be treated with caution, as there is considerable variation between subjects in the number of articles. Table 8.1 presents the data on only 53 articles in the SCI set (the citation patterns for one article, classified as 'Multidisciplinary Sciences', that was cited over 100,000 times could not be obtained). In the table, n denotes the number of articles in the category, '71-76', '77-82', '83-88', '89-94', '95-00; and ' 01-06' respectively the percentage of all citations in the period 1971 to 2006 that occur in each six year period, and 'Peak Year' the year of peak citation.

Table 8.1: Mean citation patterns of 53 of the 54 articles in the SSCI and SCI sets cited at least 1,000 times

Category	n	71-76	77-82	83-88	89-94	95-00	01-06	Peak Year
SCI	53	16.0	21.1	18.7	16.1	14.5	13.6	1985.4
Statistics & Probability	3	3.9	14.2	18.7	14.8	19.3	29.1	1998.7
Physics, Atomic, Molecular & Chemical	5	3.9	11.4	14.3	19.6	22.4	28.6	1997.4
Chemistry, Multidisciplinary	3	5.8	13.3	17.4	20.8	22.2	20.6	1994.3
Mathematics, Interdisciplinary Applications	3	5.9	15.7	17.5	14.6	19.0	27.3	1989.3
Biochemistry & Molecular Biology	7	19.5	24.6	19.1	15.8	11.9	9.2	1982.9
Physiology	3	21.0	31.1	22.1	12.4	7.3	6.1	1979.3
Multidisciplinary Sciences	9	28.1	23.5	17.1	13.3	10.8	7.2	1979.2
SSCI	33	9.6	18.2	19.0	17.2	16.7	19.3	1991.8
Economics	6	6.1	14.0	18.0	19.5	18.8	23.6	1995.0
Statistics & Probability	5	6.4	18.2	19.2	16.3	17.4	22.4	1994.0
Mathematics, Interdisciplinary Applications	4	8.6	19.3	19.1	15.3	15.8	22.0	1986.5
Psychology, Multidisciplinary	4	16.8	20.5	16.2	15.8	15.6	15.2	1983.8

Table 8.1 was used when selecting the cases, choosing from each set two subjects with mean late citation and one with mean early citation. For each case study the citation patterns of the 36 most highly cited articles published between 1969 and 1971 were investigated, with a view to: (a) Examining whether the conjectures presented in Section 8.1 apply to the cases, and (b) Identifying indicators that for the cases are more accurate predictors of total citation ranking and total citation than the standard indicator.

The decision to investigate 36 articles is empirically grounded; this research seeks to investigate a strata of highly cited articles where late citation was prevalent and found, when examining the raw data, that for some subjects (e.g., 'Physics, Atomic, Molecular & Chemical' and 'Chemistry, Multidisciplinary') the average level of late citation was considerably higher amongst the most highly cited 13 to 24 than amongst the most highly cited 25 to 36. There are two advantages in investigating articles published more than 35 years ago: (a) Patterns covering 35 years seem to reflect more closely the final patterns than those covering 20 years, and (b) An objective is to evaluate the effectiveness of predicting the ranking of total citations, and a longer period of time provides a more accurate estimate of total citations than does a shorter period. The Spearman correlation was used to evaluate the accuracy of predicting total citation ranking and the Pearson correlation between the log values of the citations to evaluate the accuracy of predicting total citation. The log values were used, as it is not assumed that the citations are normally distributed. The Pearson log test conducts the Pearson test on log values, to investigate correlations in values that are not normally distributed; normality can be tested using the Kolmogorov-Smirnov Normality tests, as implemented in Chapter 6.

Caution is advised when using Table 8.1 to compare the SCI set with the SSCI set because the sets use different criteria (at least 1,000 citations and at least 500 citations). Moreover, the SCI database grew about twice as fast as the SCI database: For the SCI 3.9 times as many documents were published in 2006 than in 1970 whereas for the SSCI the figure is only 1.9. A third problem is that 75.8% (25 out of 33) of the articles in the SSCI set are also in the SCI database; of the 8 articles in the SSCI set not on the SCI database, 5 are classified as 'Economics', 1 as 'Business, Finance', 1 as 'Business' and 'Management', and 1 as 'Psychology, Social'. There is less overlap amongst less highly cited articles: Of the 45,749 SSCI articles published in 1970, only 35.1% are also in the SCI database.

8.3: Results

In this section findings are presented relating to the six conjectures presented in Section 8.1 and indicators are identified that, for the cases examined, are more accurate predictors of total citation ranking and total citation than the standard indicator. For brevity 'Physics, Atomic, Molecular & Chemical' are denoted as 'Physics', 'Chemistry, Multidisciplinary' as 'Chemistry', 'Psychology, Multidisciplinary' as 'Psychology' and 'Statistics & Probability' as 'Statistics'.

8.3.1: Citation pattern diversity amongst highly cited articles published in 1970

This section investigates whether Conjecture 1 applies to the 54 articles in the SCI set and 33 articles in the SSCI set. The two sets overlap, in that 9 articles are in both sets. The purpose of this section is to illustrate that even a small number of articles can have patterns of annual citation more diverse than those reported in the research of Aversa (1985), Cano and Lind (1991), Aksnes (2003) and Levitt and Thelwall (2007a). Figure 8.1 presents examples of the diverse patterns of citations of the articles in the SCI set; this figure is based on Table 8.2, which also includes the total citations during the period 1971 to 2006 ('Cited 71-06') and the year of peak citation ('Peak Year').

Table 8.2: Percentage of citations in 6 six-year periods for nine highly cited SCI articles published in 1970

First author	Citations 71-06	71-76	77-82	83-88	89-94	95-00	01-06	Peak year
1. Baltimor D	1394	47.1	13.0	11.6	9.6	12.1	6.6	1972
2. Carroll JD	1001	13.7	24.8	16.1	14.6	13.8	17.1	1978
3. Breslow N	1095	3.6	21.9	35.7	22.0	12.0	4.8	1985
4. Needlema SB	2950	1.2	2.3	12.2	33.3	26.4	24.6	1993
5. Bretsche P	1019	12.9	9.9	8.7	20.2	26.9	21.4	1996
6. Boys SF	5953	.2	1.7	6.6	13.7	28.5	49.3	2006
7. Rehm D	2450	3.7	12.2	17.1	21.6	23.4	22.0	2006
8. Marshall WA *	1129	6.6	9.1	14.8	19.2	20.3	29.9	2006
9. Brune JN	1242	10.3	13.9	14.2	15.9	20.7	25.0	2006
Mean	**2026**	**11.0**	**12.1**	**15.2**	**18.9**	**20.5**	**22.3**	**1994**

* This article is also in the SSCI set

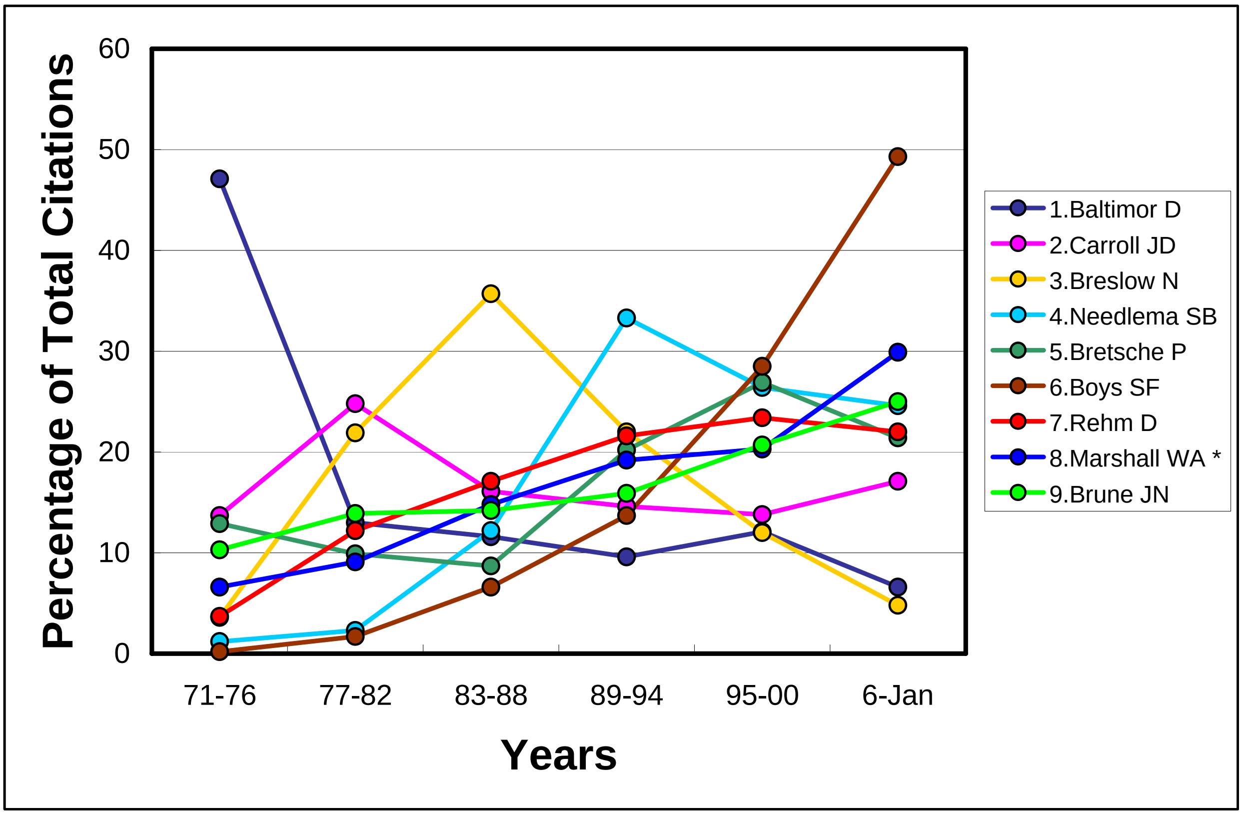

Figure 8.1: Percentage of citations in 6 six-year periods for nine highly cited SCI articles published in 1970 (*article also in the SSCI set)

The peak years of citation of the first five items in Figure 8.1 are each in different six-year periods between 1971 and 2000, whereas the citations of the final four, all peak in 2001−06. The peak is particularly steep for the first and sixth items, and particularly flat for the fifth and seventh. Amongst the four examples of peak citation in 2001–06, the citations in the first six-year period ranged from .2% to 10.3% of total citations and the citations in the first 18 years ranged from 8.5% to 38.4% of total citations.

Figure 8.2 presents examples of the diverse patterns of annual citations of the articles in the SSCI set; this figure is based on Table 8.3, which also includes the total citations during the period 1971 to 2006 and the year of peak citation. The peak years of citation of the first five items in Figure 8.2 are each in different six-year periods between 1971 and 2000, whereas the citations of each of the final four, peak in 2001–06. The peak is particularly steep for the second and sixth items and particularly flat for the third and fourth. Amongst the four examples of peak citation in 2001–06, the citations in the first six-year period ranged from .3% to 7.8% of total citations and the citations in the first 18 years ranged from 1.5% to 47.7% of total citations. Figures 8.1 and 8.2 indicate that Conjecture 1 applies to highly cited articles published in 1970, in that there is very considerable diversity in their citation patterns. Using the same notation as Table 8.2, in Table 8.3 'Cited 71-06' denotes the total citations during the period 1971 to 2006 and 'Peak Year' the year of peak citation.

Table 8.3: Percentage of citations in 6 six-year periods for nine highly cited SSCI articles published in 1970

First author	Citations 71-06	71-76	77-82	83-88	89-94	95-00	01-06	Peak year
1. Bolles RC	781	29.4	23.2	14.1	13.4	12.4	7.4	1975
2. Clewell DB	751	10.0	41.5	36.5	9.7	1.7	.5	1981
3. Tomlinso BE	1173	3.5	17.8	29.0	24.5	14.9	10.3	1985
4. Rizzo JR	1046	5.0	13.6	18.5	23.2	21.9	17.8	1992
5. Akaike H	511	3.9	18.2	16.2	19.8	23.3	18.6	1996
6. Hastings WK *	1163	.3	.4	.8	7.3	31.1	60.1	2006 & 2004
7. Brislin RW	614	3.6	4.4	6.5	11.6	24.9	49.0	2006
8. Robins E	657	6.4	12.6	12.9	17.2	19.5	31.4	2005
9. Hoerl AE *	1106	7.8	20.3	19.6	15.0	14.7	22.5	2006
Mean	**867**	**7.8**	**16.9**	**17.1**	**15.7**	**18.3**	**24**	**1995**

* This article is also in the SCI set

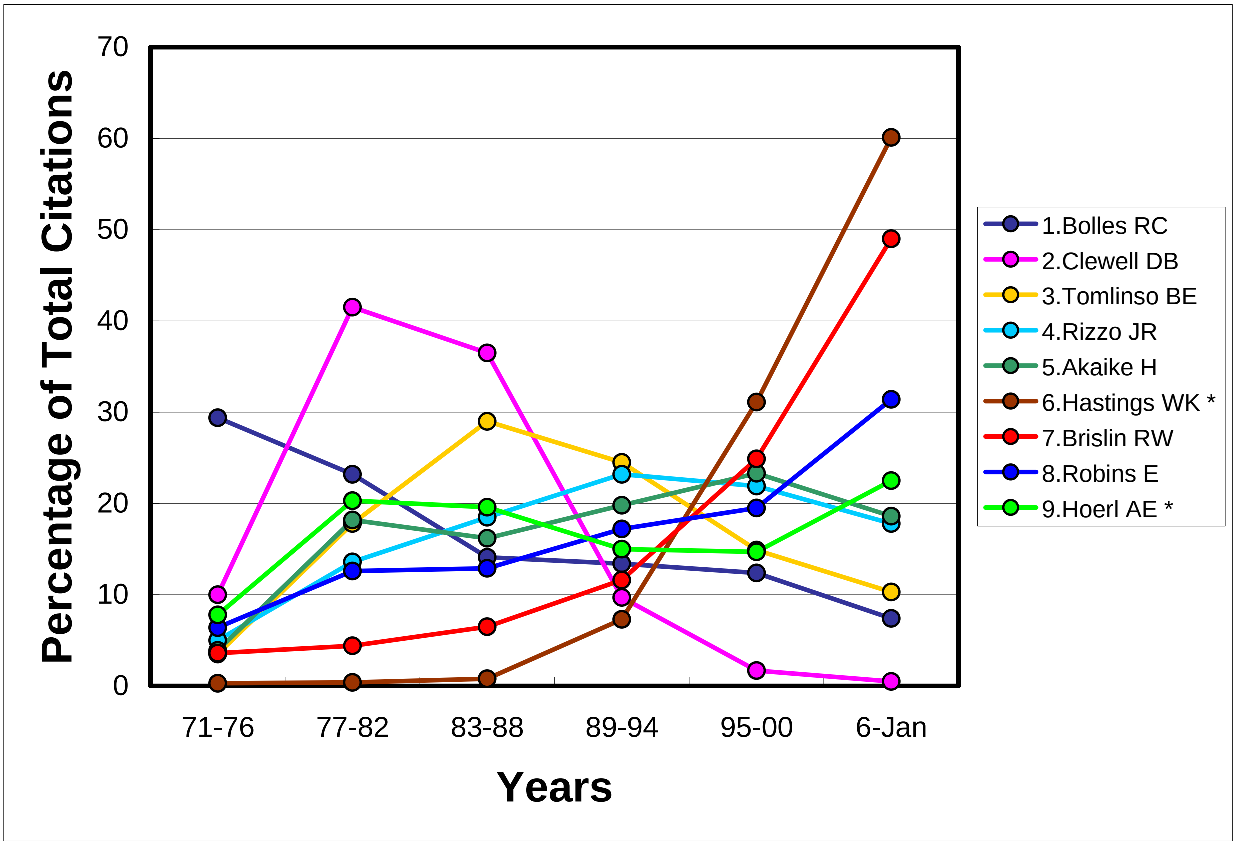

Figure 8.2: Percentage of citations in 6 six-year periods for nine highly cited SSCI articles published in 1970 (*article also in the SCI set)

8.3.2: Late citation

In order to test whether Conjecture 2 applies to the 6 cases, percentages were calculated by dividing the number of citations during the first six years by the total citations to date (data presented in Table App.3, Appendix). Table 8.4 presents the three lowest percentages (denoted by 1st, 2nd and 3rd) and the three highest percentages (denoted by 34th, 35th and 36th). In the table, 'Mean (of 1st to 36th)' denotes the average percentage of early citation of the 36 articles, 'Mean/3rd' denotes the ratio of the values of 'Mean (of 1st to 36th)' to that of '3rd' and '34th/mean' denotes the ratio of the values of 34th' to that of 'Mean (of 1st to 36th)'.

Table 8.4: Percentage of citations accumulated in the first six years for articles with particularly low or high percentages of early citation from the 36 most highly cited articles in each subject area

Subject	1st	2nd	3rd	34th	35th	36th	Mean (of 1st to 36th)	Mean /3rd	34th/ mean
Physics	.2	.7	1.5	19.4	22.2	25.7	12.6	8.4	1.5
Chemistry	.7	.9	1.4	36.9	38.0	52.8	18.6	13.3	2.0
Physiology	3.6	4.6	5.2	35.3	35.6	43.1	18.3	3.5	1.9
Economics	1.8	2.5	3.7	19.4	22.2	25.7	10.6	2.9	1.8
Statistics	.3	.7	1.7	18.8	20.1	20.2	8.3	4.9	2.3
Psychology	3.7	3.7	4.5	33.3	39.4	40.5	16.6	3.7	2.0

Table 8.4 indicates that Conjecture 2 applies to all subjects, although more strongly to Chemistry, Physics and Statistics than to Physiology, Economics and Psychology. The ratio of the mean to the 3rd varies considerably between subjects, ranging from 2.9 for Economics to 13.3 for Chemistry. The ratio of the 34th to the mean varies much less between subjects, ranging from 1.5 for Physics to 2.3 for Statistics.

8.3.3: Differences in mean pattern of citation between highly cited articles and less highly cited articles

The Spearman correlation between the ranking and percentage of citations in the first 6 years after publication is presented in Table 8.5 (using Table App.3). From Table 8.5, Conjecture 3 is validated in that for Physics and Economics the correlations are over .55 with high statistical significance, whereas for Physiology the correlation was low and not statistically significant.

Table 8.5: Spearman correlation between percentage of early citations and ranking by all citations to date

Subject	Correlation
Physics	.63**
Chemistry	.45**
Physiology	.15
Economics	.57**
Statistics	.35*
Psychology	.43**

* Significant at *p*=.05; ** Significant at *p*=.01

8.3.4: Mean citation patterns

Figures 8.3 to 8.8 present for each year from 1991 to 2006 and for four sets of articles (the 12 most highly cited articles, those cited 13th to 24th, those cited 25th to 36th, and all except the most highly cited 36) the number of citations expressed as a percentage of total citations to date.

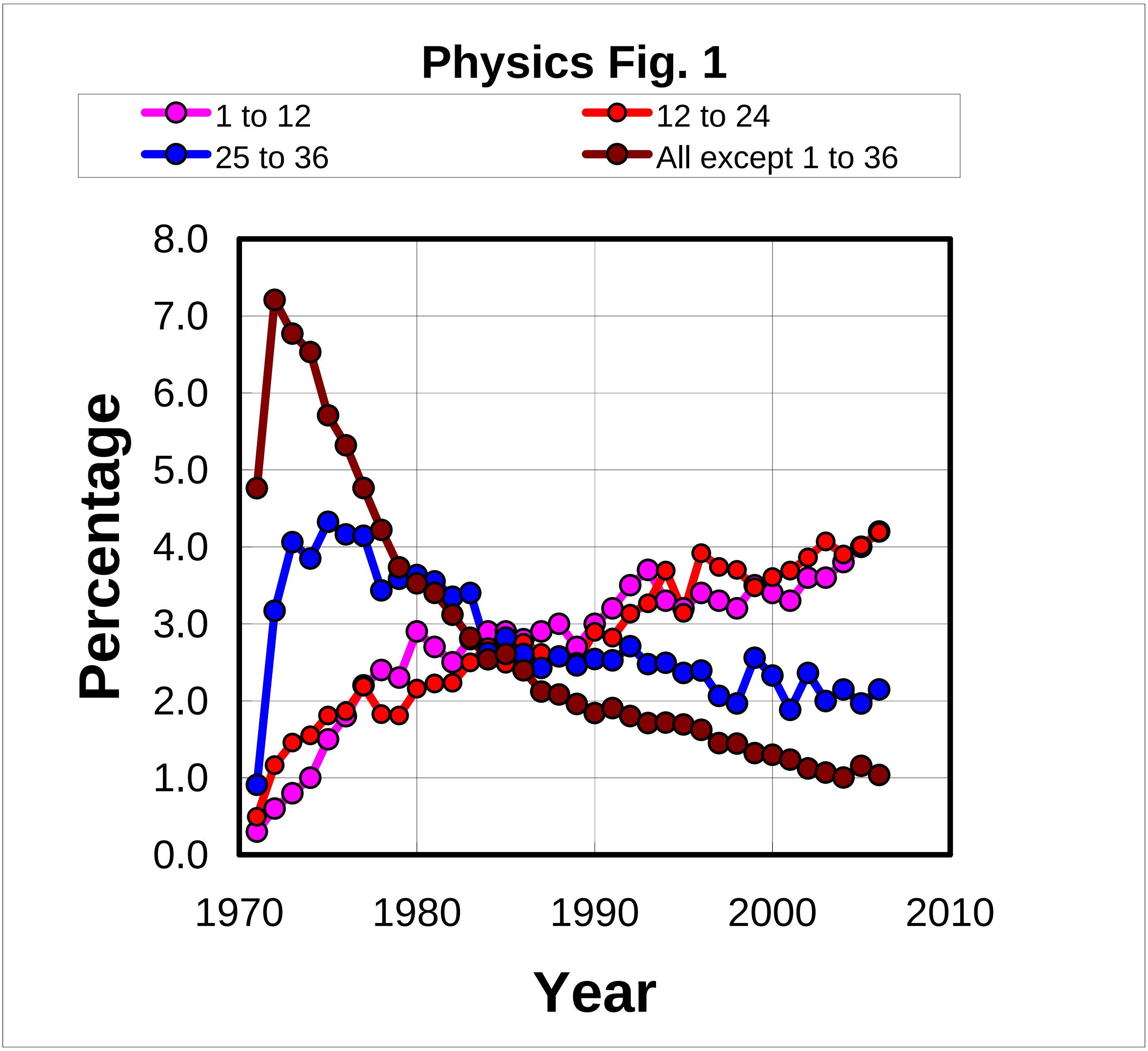

Figure 8.3: Percentage of citations accumulated over time for four sets of highly cited articles in Physics, Atomic, Molecular & Chemical SCI articles (1969–71)

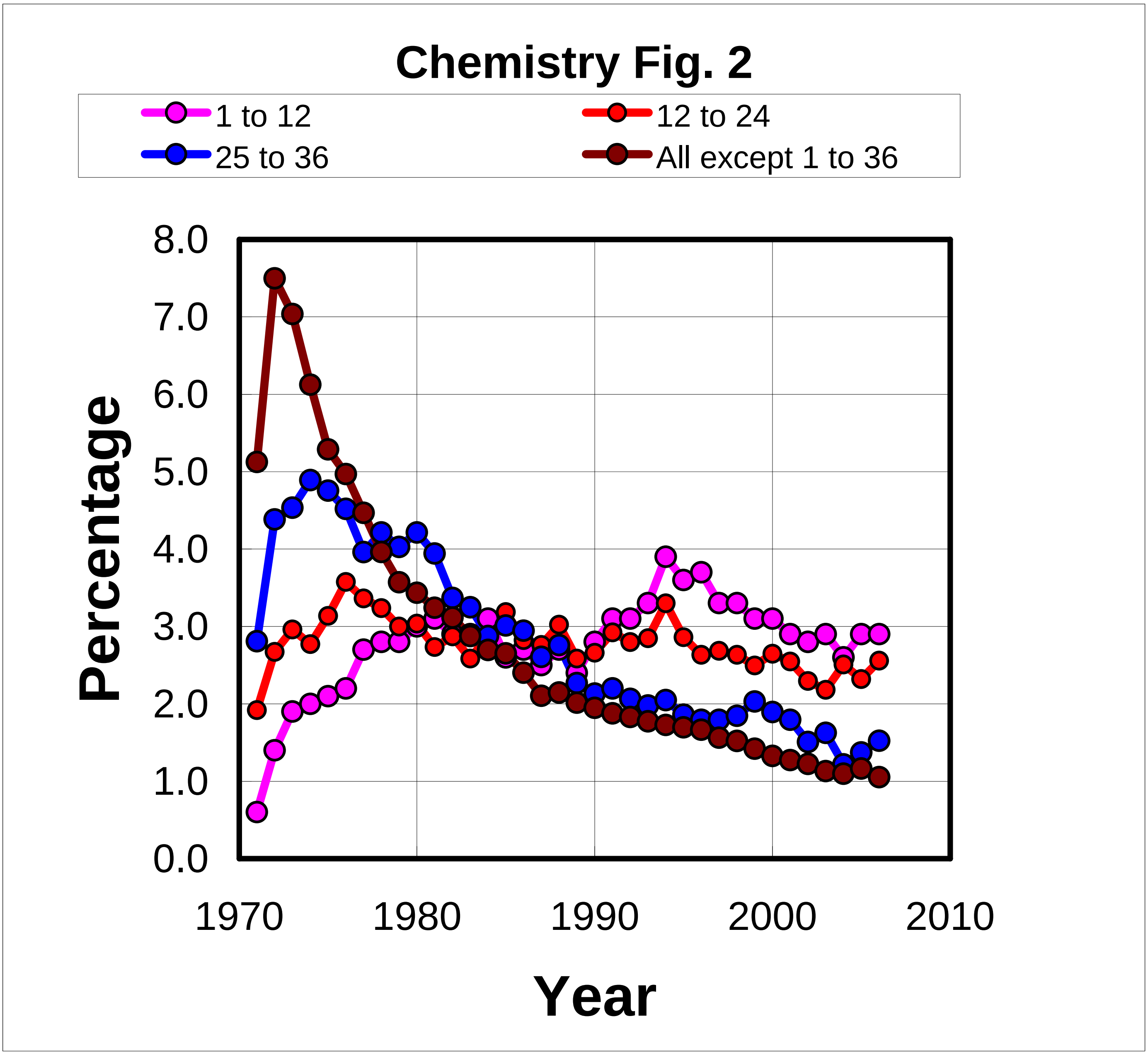

Figure 8.4: Percentage of citations accumulated over time for four sets of highly cited articles in Chemistry, Multidisciplinary SCI articles (1969−71)

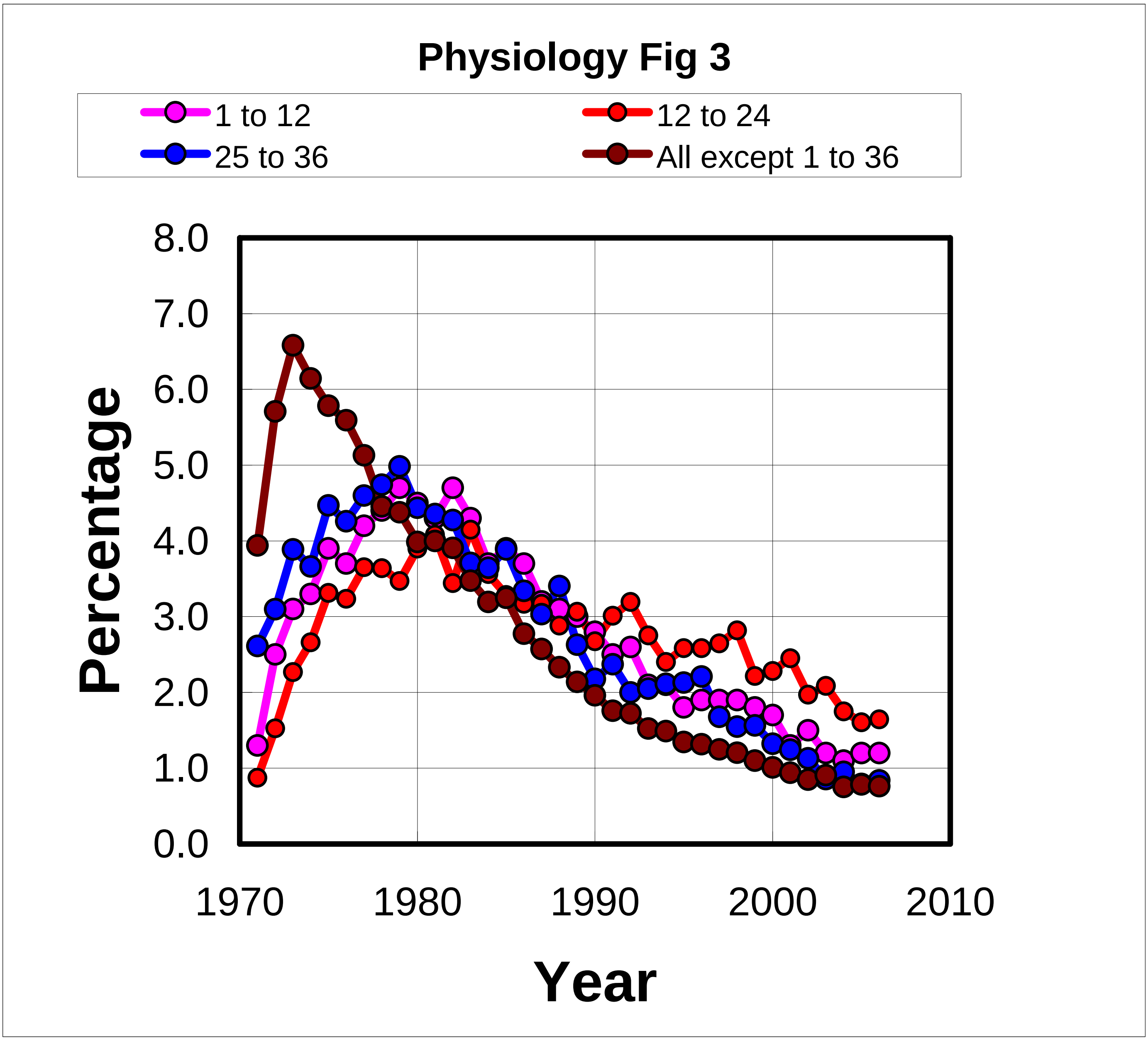

Figure 8.5: Percentage of citations accumulated over time for four sets of highly cited articles in Physiology SSCI articles (1969–71)

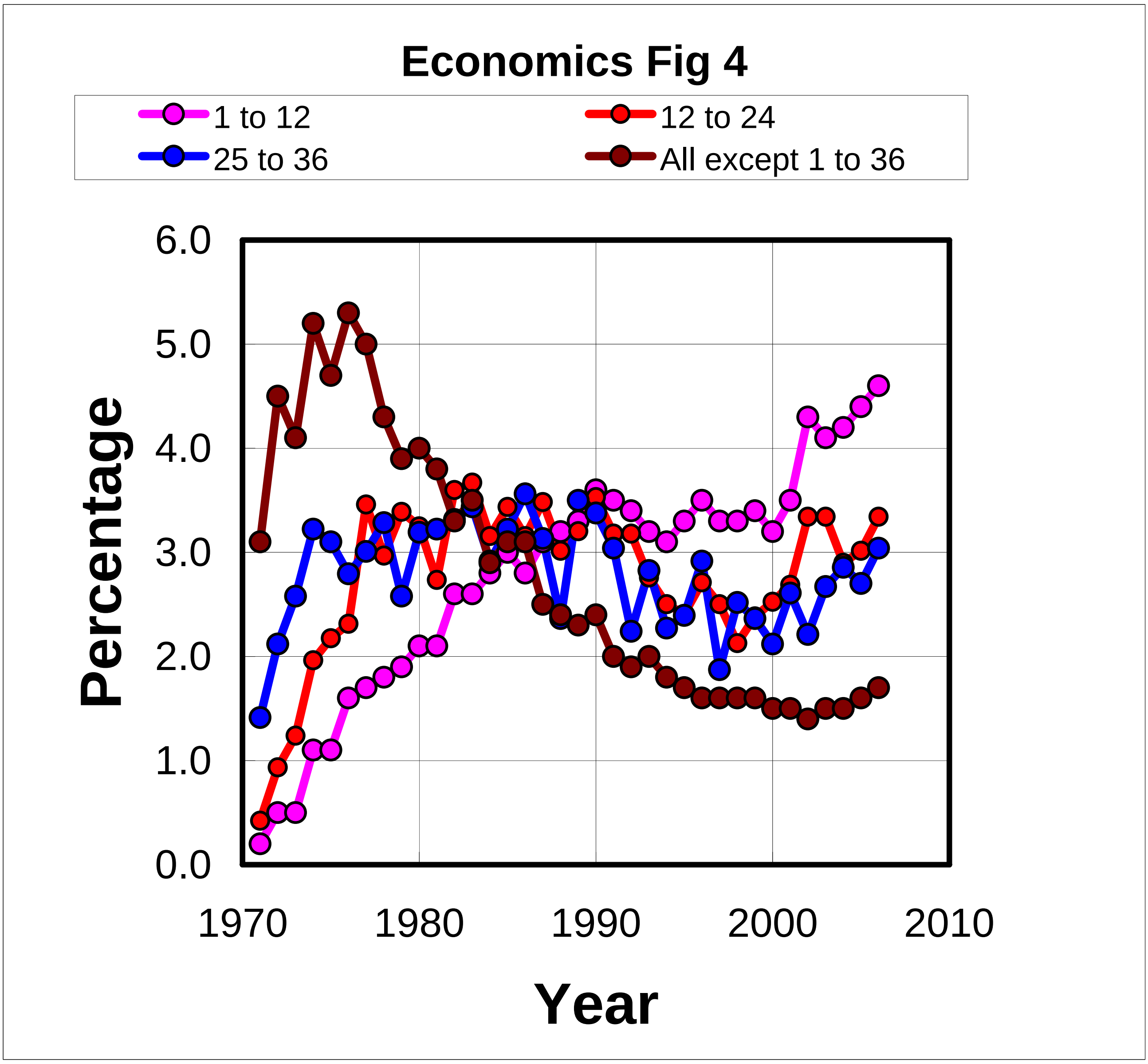

Figure 8.6: Percentage of citations accumulated over time for four sets of highly cited articles in Economics SSCI articles (1969–71)

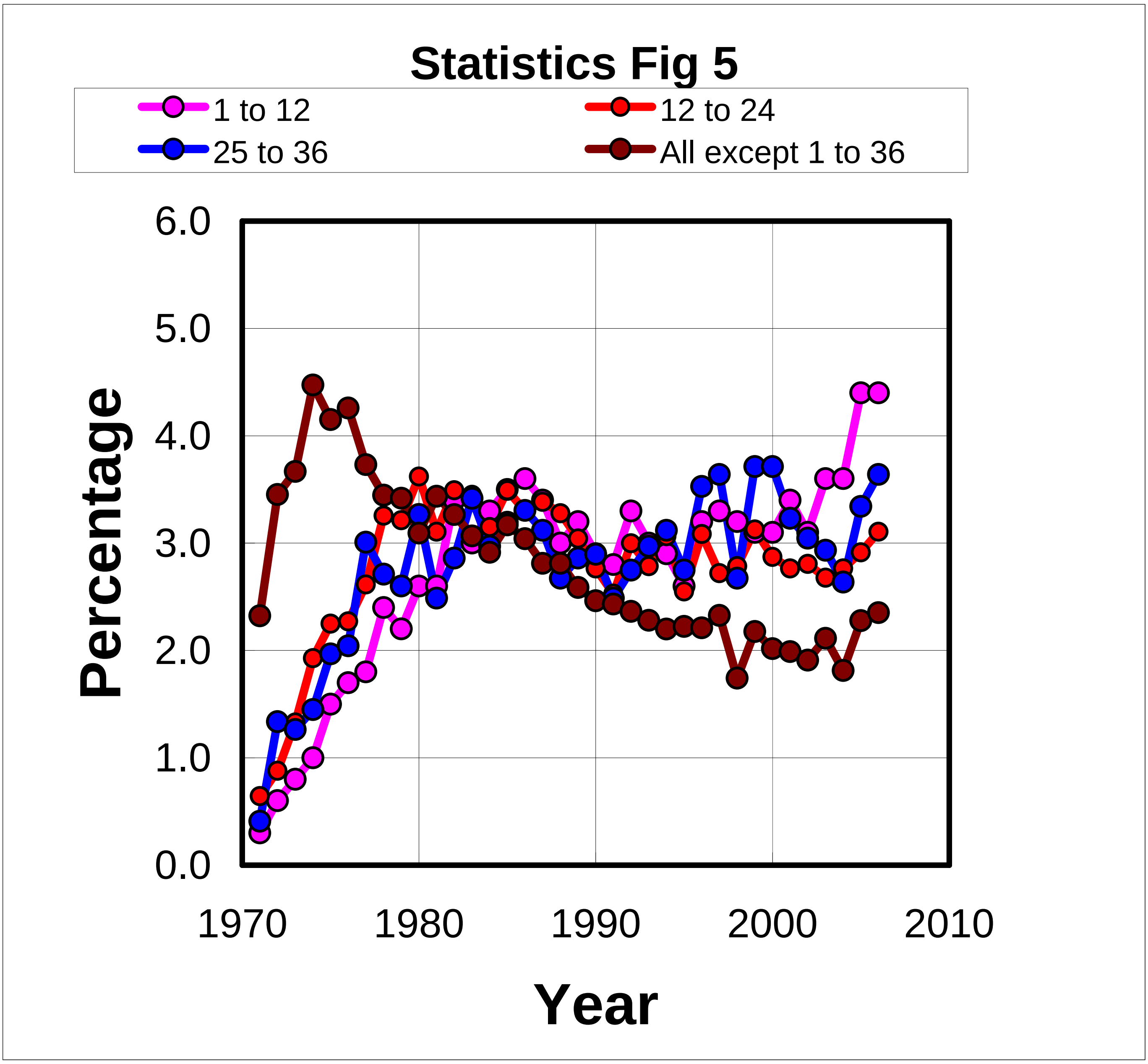

Figure 8.7: Percentage of citations accumulated over time for four sets of highly cited articles in Statistics & Probability SSCI articles (1969–71)

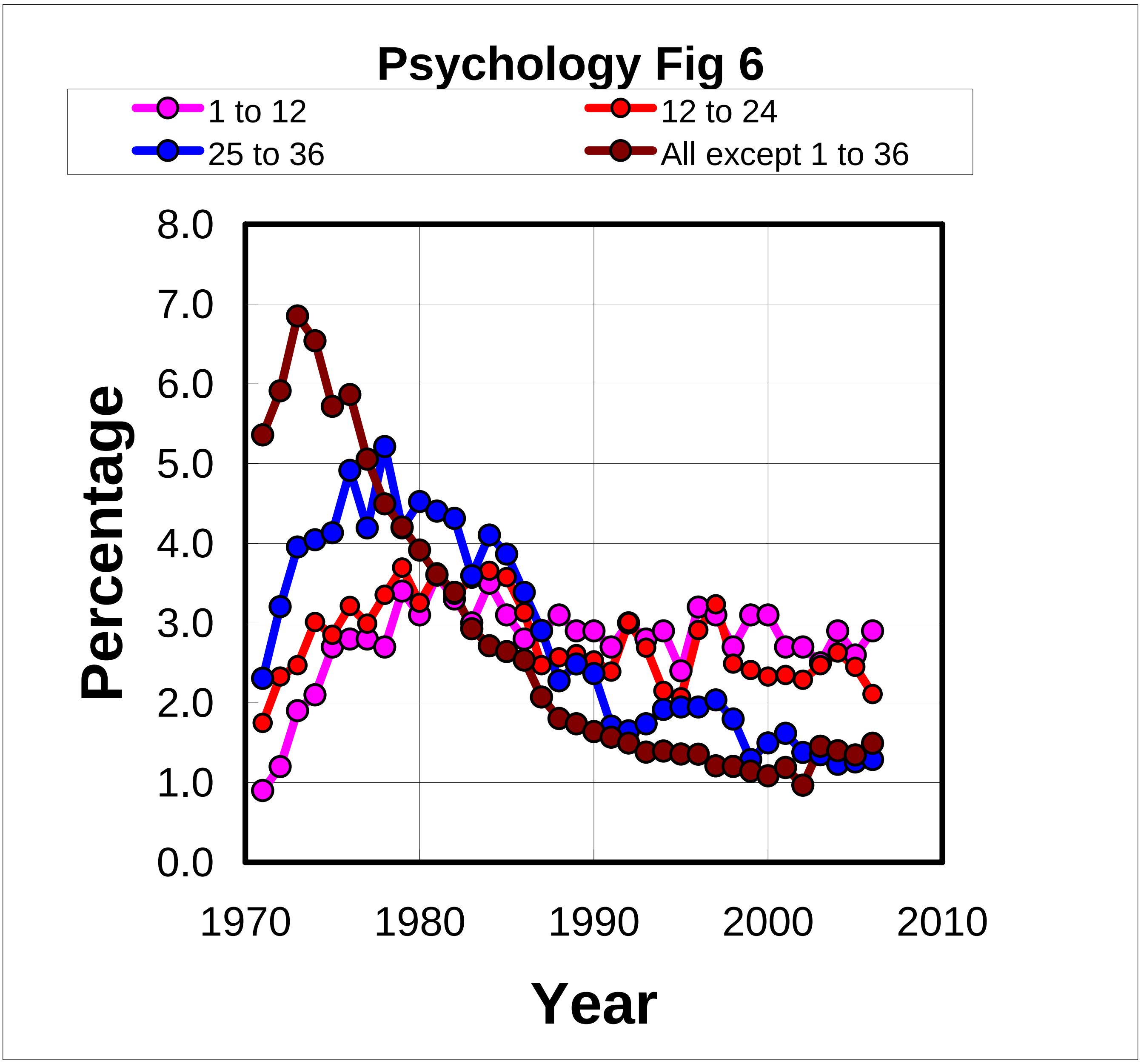

Figure 8.8: Percentage of citations accumulated over time for four sets of highly cited articles in Psychology, Multidisciplinary SSCI articles (1969−71)

Because of the small numbers in the categories ‘1 to 12’, ‘13 to 24’ and ‘25 to 36’, only broad inferences should be drawn from Figures 8.3 to 8.8. Conjecture 4 applies to some extent in that for every subject, the category ‘All except 1 to 36’ rises more sharply in the first few years than does any of the other sets. However, there are sharp differences between subjects. For instance, for Physics, Statistics and Economics the peak year for ‘1 to 12’ is 2006, whereas for Physiology the frequency of the peak year (1982) is 3.8 times that of 2006.

8.3.5: Frequency of late citation amongst highly cited articles published in 1986

This section examines whether Conjecture 5 applies to two samples of highly cited articles published in 1986, i.e. different data sets to that used in the previous four sections. The first sample considered is a subset of the set of 126 SCI articles published in 1986 that, to date (June 2007), have been cited at least 1,000 times. The SCI database lists these articles in 42 combinations of subjects, and the subset investigated is the collection of the single most highly cited articles in each of the 42 combinations. The rationale for investigating the highest cited article of each combination, as opposed to all 126 articles, is that the 126 is skewed towards certain subjects; specifically 41 of the articles are classified as ‘Multidisciplinary Sciences’, 14 as both ‘Biochemistry & Molecular Biology’ and ‘Cell Biology, and 10 as ‘Medicine, General & Internal’. Figure 8.9 presents the distribution of the peak citation years of the highest cited articles in each of the 42 combinations. The vertical axis represents the percentage of combinations for which the year of peak citation lies in the two-year interval specified in the horizontal axis.

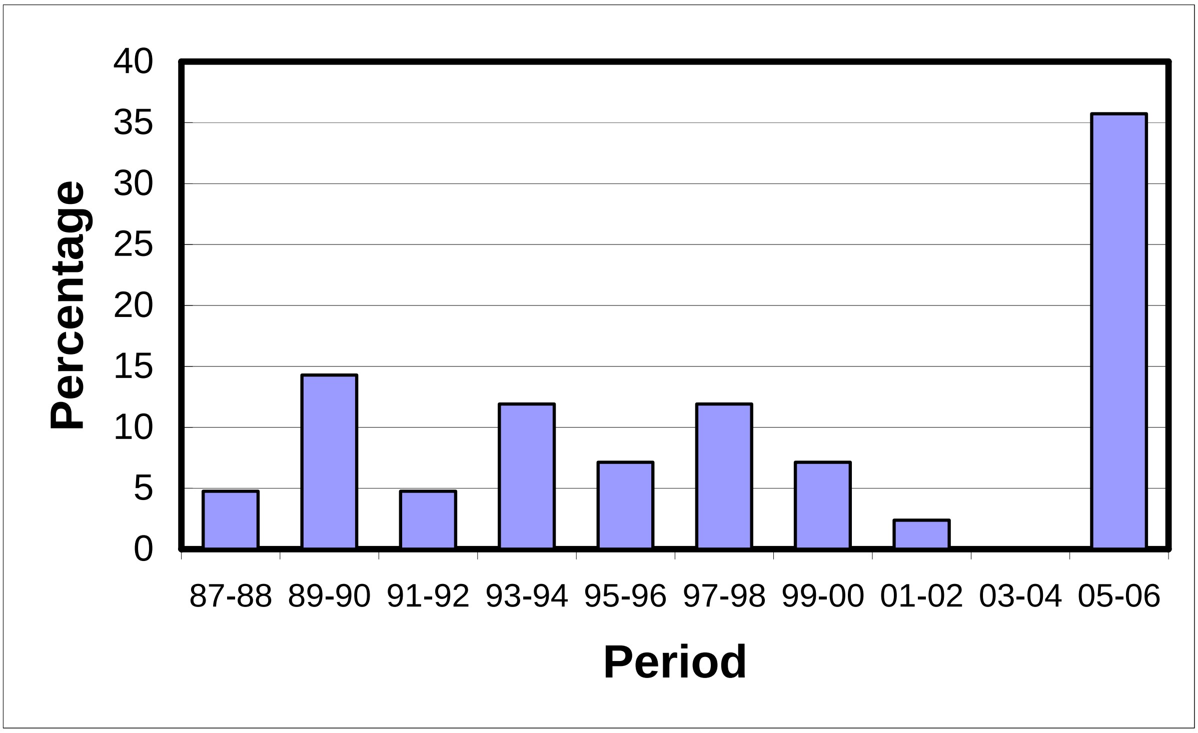

Figure 8.9: Peak citation years of the highest cited article in the 42 articles of the SCI sample

The mean peak citation year of the 42 articles was 1998.1 and for 76.2% of the articles the peak citation year was later than 1992. Table 8.6 presents the 15 categories for which the peak citation year was 2005 or 2006. There is a wide range of subject categories for which the peak year of citation of the most highly cited article is at least 19 years after publication (1986). Three of the subject combinations consisted of three subjects, 4 of two subjects, and the remaining 8 of one subject; amongst the 27 articles with peaks prior to 2005, 3 consisted of three subjects, 6 of two subjects, and 18 of one subject. Possibly subject combinations consisting of more than one subject are more likely to experience a late citation peak.

Table 8.6: Subject combinations in the first sample for which the peak citation year of the highest cited article was 2005 or 2006

Subject combination	Year
Biology; Mathematics, Interdisciplinary Applications; Statistics & Probability	2005
Cardiac & Cardiovascular Systems	2005
Chemistry, Analytical	2005
Chemistry, Physical	2005
Medicine, General & Internal	2005
Physiology; Sport Sciences	2005
Biochemical Research Methods; Immunology	2006
Biochemistry & Molecular Biology; Evolutionary Biology; Genetics & Heredity	2006
Computer Science, Artificial Intelligence; Engineering, Electrical & Electronic	2006
Ecology	2006
Economics; Mathematics, Interdisciplinary Applications; Social Sciences, Mathematical Methods	2006
Geochemistry & Geophysics	2006
Medicine, Research & Experimental; Pharmacology & Pharmacy	2006
Physics, Multidisciplinary	2006
Rheumatology	2006

The second sample considered consists of the most highly cited five articles published in 1986 in 13 different SSCI subjects. It was decided not to use a similar criterion to that used to obtain the first sample, as a sample obtained in that way would have been very strongly biased towards subjects regarded as science and psychology. The 13 subjects were chosen on the basis that excluding duplication (such as many varieties of psychology) they were the subject categories in which most SSCI articles were published in 1986. For each category the peak citation year of the five most highly cited articles was obtained. Figure 8.10 presents the distribution of these peak citation years. As in Figure 8.9, the vertical axis represents the percentage of combinations for which the year of peak citation lies in the two-year interval specified in the horizontal axis.

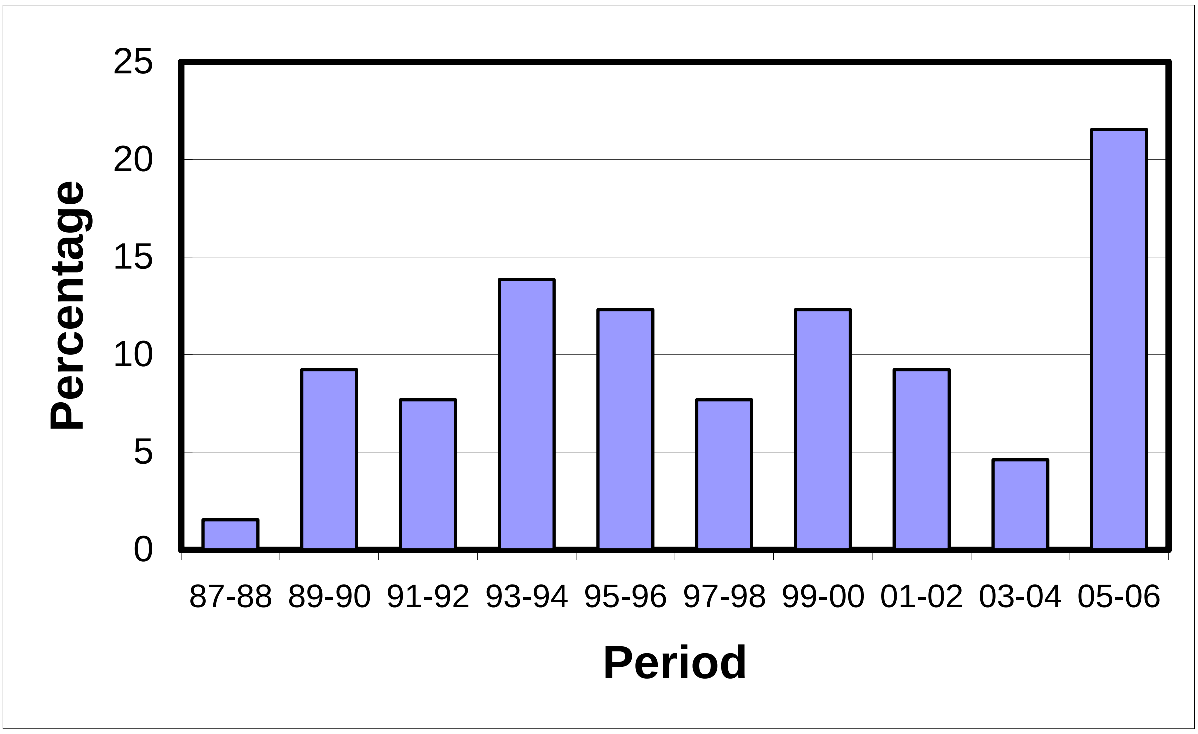

Figure 8.10: Peak citation years for the 65 articles in the SSCI sample

The mean peak citation year of the 65 articles was 1998.0 and for 81.5% of the articles the peak citation year was later than 1992. Table 8.7 presents the mean peak citation years for the 13 categories. Seven of the thirteen subjects in Table 8.7 have mean citation peaks more than 12 years after 1986. For eight subjects (Psychology, Economics, Planning & Development, Sociology, Political Science, Education & Educational Research, International Relations and Information Science & Library Science) the range of peak citation years for the 1st to 5th articles was more than ten years. In the table, '1st' to '5th' denote the years of peak citation for the 1st to 5th most highly cited articles and 'Mean' the average of peak citation years of the five most highly cited articles.

Table 8.7: Peak citation years for the articles in the second sample

Subject	1st	2nd	3rd	4th	5th	Mean
Management	2006	2000	2001	2005	2005	2003.4
Psychology	2006	1994	2006	2005	2001	2002.4
Economics	1997.5	2006	2006	2006	1995	2002.1
Planning & Development	2003	2006	1993	1997	2001	2000.0
Sociology	2006	2000	2006	1994	1993.5	1999.9
Political Science	1999.5	2006	1990	2002	2000.5	1999.6
Education & Educational Research	1996	1993	1999	1999	2004	1998.2
Public, Environmental & Occupational Health	1999	1997	1999	1995	1996	1997.2
International Relations	1998	1992.7	2004	1994.5	1991.5	1996.1
Psychiatry	1995	2000	1996	1996	1991.5	1995.7
Social Sciences, Interdisciplinary	1995	1993	2002	1998	1988	1995.2
Information Science & Library Science	1993	1993	1989	2002	1989	1993.2
Law	1990	1989	1990	1991	1992	1990.4
Mean	**1998.8**	**1997.7**	**1998.5**	**1998.8**	**1996.0**	**1998.0**

Conjecture 5 applies to both samples, in that late citation peaks are present amongst a wide diversity of subjects or subject combinations. In addition, for both samples over 75% of the articles have their citation peaks more than six years after publication and the mean peak year of annual citation is approximately 12 years after publication.

8.3.6: Predicting total citation ranking and total citations

This section first examines Conjecture 6 and then seeks to identify an indicator that, for the cases, is a more accurate predictor of total citation ranking and total citation than the standard indicator. Tables 8.8 and 8.9 compare the extent to which two indicators correlate with citation to date: The standard indicator (the sum of citations for the years after publication) and the alternative indicator (the sum of the citations for the final two years). The second indicator was chosen on the basis of finding that for some of the cases there is a correlation between early citation and citation ranking (Section 8.3.3). Table 8.8 compares the extent to which total citation to date correlates with the sum of citations during the first six years after publication ('Spearman rank sum 1 to 6' and 'Pearson log sum 1 to 6') with the extent to which total citation to date correlates with the sum of citations during the fifth and sixth years after publication '(Spearman rank sum 5 and 6' and 'Pearson log sum 5 and 6'). In the table, 'Spearman rank' demotes the p-value of the Spearman test and 'Pearson log' the p-value of the Pearson test on the logarithmic values.

Table 8.8: Correlation between indicators based on citations in the first six years after publication and the citation ranking and citations to date

Subject	**Spearman rank sum 1 to 6**	**Spearman rank sum 5 and 6**	**Pearson log sum 1 to 6**	**Pearson log sum 5 and 6**
Physics	.07	.16	.13	.26
Chemistry	.06	.18	-.05	.01
Physiology	.37 *	.45 **	.40 *	.49 **
Economics	.17	.45 **	.11	.37 *
Statistics	.23	.35 *	.15	.26
Psychology	.19	.37 *	.21	.37 *

* Significant at p=.05; ** Significant at p=.01

As in the previous table, Table 8.9 compares the extent to which total citation to date correlates with the sum of citations during the first ten years after publication ('Spearman rank sum 1 to 10' and 'Pearson log sum 1 to 10') with the extent to which total citation to date correlates with the sum of citations during the ninth and tenth years after publication '(Spearman rank sum 9 and 10' and 'Pearson log sum 9 and 10'). Again, 'Spearman rank' demotes the p-value of the Spearman test and 'Pearson log' the p-value of the Pearson test on the logarithmic values.

Table 8.9: Correlation between indicators based on citations in the first ten years after publication and the citation ranking and citations to date

Subject	Spearman rank sum 1 to 10	Spearman rank sum 9 and 10	Pearson log sum 1 to 10	Pearson log sum 9 and 10
Physics	.23	.11	.41 *	.26
Chemistry	.25	.42 *	.10	.39 *
Physiology	.44 **	.51 **	.51 **	.65 **
Economics	.45 **	.64 **	.37 *	.57 **
Statistics	.33 *	.34 *	.21	.23
Psychology	.32	.53 **	.37 *	.50 **

* Significant at p=.05; ** Significant at p=.01

Tables 8.8 and 8.9 confirm Conjecture 6, in that for Physiology all eight correlations were statistically significant whereas for Physics only one of the correlations was statistically significant. In Table 8.8, both indicators provide statistically significant results for Physiology, but the correlation is stronger when using the alternative indicator. In addition, the standard indicator provides no other statistically significant correlations, whereas the alternative indicator provides five other statistically significant correlations. In Table 8.9, both indicators provide statistically significant correlations for Economics, Physiology, and Psychology but in each case the correlation produced by the alternative indicator is stronger. In addition, only in one instance (the Pearson log for Physics) is the correlation with the alterative indicator of lower statistical significance than the correlation with the standard indicator.

8.4: Discussion

The six conjectures apply to the subjects, periods, and strata of highly cited articles investigated. Although it seems likely that the conjectures apply to many subjects and time periods, future research on other subjects and different time periods would be needed to confirm this because the subjects studied were not chosen at random.

On the basis of Table 8.5 and Figures 8.3 to 8.7, it seems that whether the conjectures apply to a subject and a period depends critically on the strata of highly cited articles. As mentioned above, by choosing the most highly cited 36 articles in a subject in a three-year period, a stratum of highly cited articles where late citation was prevalent was deliberately selected. It is anticipated that the conjectures would have applied less strongly if, for instance, the most highly cited 72 articles in a subject in a three-year period had been considered.

Although the obliteration phenomenon proposed by Merton (1968) and Garfield (1975, 1993) may have reduced the level of citations of some of the articles we examined, many of the articles we examined were cited late. The high frequencies of late citation discussed above contrast with the low frequencies found by Glänzel, Schlemmer and Thijs (2003) and Van Raan (2004). In line with the findings, the differences are probably due to the late citation/highly cited criteria being less stringent/more stringent respectively. Adams (2005) found that the most highly ranked papers initially will remain amongst the higher ranked papers on average; in the present chapter citation ranking in the first six years was not found to be strongly correlated with total citation ranking. One possible reason for the differences between these findings is that the criteria for an article to be classified as 'highly cited' are more stringent than those used by Adams.

8.5: Concluding remarks

The conjectures were found to apply to the six samples of articles published in 1969 to 1971 and two samples of articles published in 1986. The results indicate the presence of a number of characteristics of highly cited articles.

1. Considerable variation between articles in the patterns of citation of highly cited articles.
2. Considerable variation between articles in the level of early citation expressed as a percentage of total citation.
3. Considerable variation in the correlation between the percentage of early citations and total citation ranking.
4. Substantial differences between the mean citation patterns of the highly cited articles and those of other articles.
5. Late peaks in citation are not rare (recall that late peaks may be associated with fruitful, broadly relevant theoretical insights or broadly useful methodologies).
6. Considerable variation in the level of accuracy with which total citation ranking and total citations can be predicted from early citation levels.

This chapter suggests that investigations of highly cited articles look for the presence of these characteristics and, when they are present, take them into account. The findings on Conjectures 3 and 4 indicate that the presence of these characteristics depends on the level of citation and for this reason these characteristics could be less marked amongst highly cited articles selected using criteria different from those applied in this research.

Total citation prediction appears to be more accurate if based upon citations in years 5 and 6 only, rather than on the sum of the citations in years 1-6, at least for articles that are eventually highly cited. This alternative indicator is recommended as part of a flexible approach where the method used for predicting total citations depends on the citation patterns of the subject. The standard indicator seems well suited to predicting in the case of articles that have most of their citations during the first few years after publication. However, in cases of widely differing citation patterns, this chapter suggests using an alternative indicator that takes into account differences in citation patterns.

The fourth global research question asks whether, in general, very highly cited articles have different patterns of annual citation to less highly cited articles. In Chapter 8 the main findings are: (a) Four of the six subjects investigated have a correlation of over .42 between the percentage of early citations and total citation ranking with more highly ranked articles having a lower percentage of early citations, and (b) For all six subjects the prediction of the citation ranking of highly cited articles from the sum of citations during the first six years after publication is less accurate than prediction using the sum of their citations for only the fifth and sixth year.

Finding (a) (similar to finding (c) in the Concluding Remarks of Chapter 5) indicates that high quality research in both science and social science is prone to late citation; a likely ramification is that if research quality is judged by citation counts alone, unless a very long citation window is used, very high quality research is particularly likely to be under-valued. Finding (b) indicates that the sum of citations for the first few years after publication is not necessarily the most accurate way of predicting eventual high citation; a ramification is that high quality research might be more effectively identified not only from the total number of citations during the first few years after publication, but also from the citation trends during that period.

Chapter 9: Limitations

Although the results give clear answers to the research questions of Chapters 5 to 8, there are some limitations concerning the comprehensiveness of the data sources used, the appropriateness of the categories used and the accuracy of the techniques employed.

Some limitations apply to specific investigations and have already been discussed in the relevant chapters. For example, the incidences and h-indexes in Tables 5.1 and 5.2 that are based on fewer than 5,000 articles need to be interpreted with caution, and the h-index itself is not a reliable indicator on its own (Bornmann & Daniel, 2007). The h-indexes in Table App.2 do not take into account that some researchers are not the first authors of a high percentage of the documents that contribute to their h-index (e.g., Benbasat has an h-index of 17 in IS&LS, but this figure would be 5 if the h-index were evaluated solely on IS&LS documents in which Benbasat is the first author). Furthermore, when comparing the h-indexes of the authors, different authors may have specialised in different subject combinations that have widely differing citation behaviour. Although the sharp rise in recent citation is interesting, these findings are on a small number of articles and may not be typical of subjects other than IS&LS.

Other limitations are general, in that they apply to more than one investigation.

One general limitation is that this research investigates specific databases, document types, subjects and periods. Specifically:

- The findings are for WoS and Scopus; it is possible that the findings would be different for Google Scholar. For example: Late citation is likely to be particularly prevalent in Google Scholar as the frequency of Google Scholar citations has grown very rapidly since the inception of the Internet.
- The findings are for articles; it is possible that the findings would be different were other types of documents to be investigated. For example, because reviews are often used for reference purposes, late citation seems likely to be particularly prevalent amongst reviews.
- The findings in Chapters 5 and 6 are on a single social science subject. The considerable differences found between subjects in Chapters 7 and 8 indicate that the findings on the first authors of most highly cited articles and on collaboration are unlikely to apply to all subjects or even to all social science subjects.
- Many of the findings are for specific years. For instance, in Chapter 7 the findings for WoS are for diverse social science subjects and for 1986 and 1995, but they may have been different for other WoS science subjects or for other years. In particular, although it seems likely that the figures will not change significantly from year to year, this has not been proven. Similarly, further research is needed to establish whether the findings for Scopus for 1995 are typical of other years.

A second general limitation is that this research uses specific classification systems, namely the WoS and Scopus classification of subjects. Specifically:

- Both WoS and Scopus designate subjects at the journal level in that, for a given journal, all articles are given the same subject designation. This designation results in a coarse-grained definition of disciplinarity; articles are in a subject, if and only if, they are in journals designated to that specific subject category. A consequence of this coarse-grain is that the subject content of many of the articles may not mirror that of the journal designation. For example, one could query the extent to which articles in IRE Transactions on Information Theory are in IS&LS, as only 2.5% of the documents citing articles in that journal are in IS&LS.
- Subjects are delineated according to the indexing decisions of the database. With regard to Question 2, the indexing criteria as to which journals to include within a WoS subject may have changed between 1986 and 1995; it is not possible to quantify changes in indexing criteria, as differences between indexing criteria do not necessarily result in changed coverage between 1986 and 1995 and changed coverage does not necessarily imply different indexing criteria.
- The ISI and Scopus subject category designations may differ very substantially from how experts in the subjects would now classify the articles and the nature of the ISI categories has not varied considerably with time, although subjects and disciplines continually evolve. Whilst it would be interesting to investigate for several ISI subjects the extent to which the content of the journal matches the ISI classification, such an investigation is beyond the scope of this Ph.D.

- Both WoS and Scopus designate subjects at the journal level in that, for a given journal, all articles are given the same subject designation. This designation results in a coarse-grained definition of multi-disciplinarity; articles are multi-disciplinary, if and only if, they are in journals designated to more than one subject category. One consequence of this coarse-grain is that the subject content of many of the articles may not mirror that of the journal designation and comparison between fields with different citation levels needs to be treated with care. Whilst it would be interesting to know the extent to which journals classified in two subjects contain articles in both component subjects, this would require very extensive further research and is beyond the scope of this Ph.D.; in addition the classification of subject content by contemporary experts might not be a reliable guide to the subject content of articles published more than 20 years ago. It is possible that the findings would be different with alternative criteria for multi-disciplinarity, for instance, reference analysis (Glänzel, Schubert, & Czerwon, 1999)
- This research delineates subjects using the subject categories of WoS online rather than that of the Dialog interface. It seems likely that the findings would not be radically different were Dialog to have been used; Cronin and Meho (2006) found that the ranking of influential information scientists using the Dialog interface was little different from that using the WoS interface.

A third general limitation is that citations do not necessarily reflect quality. For example: (a) It seems likely that journals that are more widely subscribed to by institutions are more likely to be cited, and (b) Some applied research is less likely to be cited (Borgman & Furner, 2002) and in some applied fields (e.g., reciprocating internal combustion engines) the highest impact journals are not necessarily the most widely regarded and conferences are seen as the important outlets for research (Aleixandre, Valderrama, Desantes, & Torregrosa, 2004).

Chapter 10: Conclusion and further research

10.1: Conclusion

Returning to the aim and objectives of this thesis presented in Section 1.2, the aim of this thesis is to evaluate the extent to which the new facilities of WoS can be used to enhance the understanding of scholarly communication. This thesis demonstrates that these new facilities can be used to make new findings, at subject and journal levels, on issues central to scholarly communication such as collaboration and disciplinarity. Although findings at the subject-level, and to a lesser extent journal-level, are coarse grained, findings such as those in this thesis can contribute to the understanding of scholarly communication and can also be relevant to research policy.

The first primary objective is to use the new facilities on WoS online to make findings in citation analysis of interest to the field of scholarly communication; the set of findings presented in Chapters 5 to 8 and summarised Section 9.1 seem to be of interest to scholarly communication. The second primary objective is to identify some of the opportunities to use the new facilities to investigate scholarly communication and to identify some of the more interesting potential applications; Chapters 5 to 8 have used the new facilities on diverse research questions and seem to have identified some interesting potential applications. The third primary objective is to identify some of the limitations of using WoS to investigate scholarly communication; Section 9.2

summarizes some of these limitations. The secondary objective is to identify and address some of the methodological issues associated with using WoS to study scholarly communication; Chapter 4 identifies and addresses some of these methodological issues. The contribution of this thesis, described in Section 1.5, is to produce several findings about scholarly communication, to introduce and use new methods and indicators that could be used in other studies, and to identify and address some of the technical problems in using the new WoS facilities to investigate scholarly communication.

10.2: Further research

Whilst conducting this Doctoral research a number of other investigations were pursed but, in the interest of cohesion, not included in this thesis. For example: (a) Levitt and Thelwall (2008a) use the distribution of the percentages of citing documents in IS&LS to gauge the effectiveness with which WoS delineates the subject category on IS&LS, and (b) Levitt and Thelwall (2008b) investigate late citation in the Scopus subject of Engineering.

Further research related to Chapter 5 could address the following questions: (a) Is the low percentage of the highest cited articles solely in IS&LS mirrored in other subjects?, (b) How typical of WoS subjects is the correlation in IS&LS between citation ranking and lateness of citation?, and(c) How widespread is it for a journal classified in a subject to have a low percentage of citing documents in that subject and to what extent does this need to be taken into account when investigating a WoS subject? Chapter 5 uses techniques that seem to have a wider application, for example: (a) The technique of delineating all the documents in a subject can be used in investigations of subjects other than IS&LS, and (b) Another technique of interest is the examination of citation profiles (as conducted in Table 5.3); citation profiles seem to provide information in addition to the h-index that can readily be obtained using the current WoS facilities.

Similar investigations of collaboration to those in Chapter 6 could be conducted on other subject categories, time periods and researchers. Chapter 6 also introduces and uses the average partner score; this indicator of level of collaboration can potentially be used in other investigations of collaboration.

The methods of Chapter 7 can be used to investigate exceptions. For example, the combined use of subject-level and journal-level investigation can be applied to other subject categories. Additional further investigations include examining subjects for which the behaviour for 1986 is different from that for 1995 or subjects for which the behaviour for Scopus is different from that of WoS. The association found between statistical significance at the journal level and high ratios of Mean Multi to Mean Mono indicates that these ratios can be used to identify subjects that are more likely to have statistically significant correlations between citation and disciplinarity at the journal level. Chapter 7 uses two techniques that could be applied more widely: (a) It compares the levels of citation of multi-disciplinary articles and mono-disciplinary articles, and (b) It compares citation levels using the Normalised Hirsch Index (h_{norm}).

The methods used in Chapter 8 can be applied to examine patterns of annual citation for additional subjects and time periods. In Chapter 8, total citation of articles that are eventually very highly cited correlates more strongly with the sum of citations in the 5^{th} and 6^{th} years after publication than it does with the sum of citation in the 1^{st} to 6^{th} years after publication (the standard indicator). When comparing articles with widely differing patterns of annual citation, indicators that take into account differences in citation trends may provide more accurate predictions of total citation than predictions obtained using the standard indicator.

Appendix

Table App.1 (referred to in Chapter 5) presents, in decreasing order of citation ranking, the 77 most highly cited articles in IS&LS. In the table, 'Author(s)' denotes the author(s) of the article and 'Title' its title.

Table App.1: The 77 most highly cited articles in IS&LS (in the order of citation ranking)

Author(s)	**Title**
1. Davis FD (1989)	Perceived Usefulness, Perceived Ease Of Use, And User Acceptance Of Information Technology
2. Hu M (1962)	Visual-Pattern Recognition By Moment Invariants
3. Gruber TR (1993)	A Translation Approach To Portable Ontology Specifications
4. Deerwester S, Dumais ST, Furnas GW, Landauer TK and Harshman R (1990)	Indexing By Latent Semantic Analysis
5. Porter MF (1980)	An Algorithm For Suffix Stripping
6. Salton G and Buckley C (1988)	Term-Weighting Approaches In Automatic Text Retrieval
7. Max J (1960)	Quantizing For Minimum Distortion
8. Robertson SE and Jones KS (1976)	Relevance Weighting Of Search Terms
9. Taylor S and Todd PA (1995)	Understanding Information Technology Usage - A Test Of Competing Models
10. Small H (1973)	Cocitation In Scientific Literature - New Measure Of Relationship Between 2 Documents
11. Gallager RG (1962)	Low-Density Parity-Check Codes
12. Doll WJ and Torkzadeh G (1988)	The Measurement Of End-User Computing Satisfaction
13. Bates DW, Teich JM, Lee J, Seger D, Kuperman GJ, Ma'Luf N, Boyle D and Leape L (1999)	The Impact Of Computerized Physician Order Entry On Medication Error Prevention
14. Adams DA, Nelson RR and Todd PA (1992)	Perceived Usefulness, Ease Of Use, And Usage Of Information Technology - A Replication
15. Salton G and Buckley C (1990)	Improving Retrieval Performance By Relevance Feedback
16. Belkin NJ, Oddy RN and Brooks HM (1982)	Ask For Information-Retrieval .1. Background And Theory
17. Price DJD (1976)	General Theory Of Bibliometric And Other Cumulative Advantage Processes

18. Taylor RS (1968)	Question-Negotiation And Information Seeking In Libraries
19. Haynes RB, Wilczynski N, Mckibbon KA, Walker CJ and Sinclair JC (1994)	Developing Optimal Search Strategies For Detecting Clinically Sound Studies In Medline
20. Egenhofer MJ and Franzosa RD (1991)	Point-Set Topological Spatial Relations
21. Compeau DR and Higgins CA (1995)	Computer Self-Efficacy - Development Of A Measure And Initial Test
22. Kuhlthau CC (1991)	Inside The Search Process - Information Seeking From The Users Perspective
23. Daft RL, Lengel RH and Trevino LK (1987)	Message Equivocality, Media Selection, And Manager Performance - Implications For Information-Systems
24. Straub DW (1989)	Validating Instruments In Mis Research
25. Dennis AR, George JF, Jessup LM, Nunamaker JF and Vogel DR (1988)	Information Technology To Support Electronic Meetings
26. Bates MJ (1989)	The Design Of Browsing And Berrypicking Techniques For The Online Search Interface
27. Jansen BJ, Spink A and Saracevic T (2000)	Real Life, Real Users, And Real Needs: A Study And Analysis Of User Queries On The Web
28. Stockwell D and Peters D (1999)	The Garp Modelling System: Problems And Solutions To Automated Spatial Prediction
29. Venkatesh V, Morris MG, Davis GB and Davis FD (2003)	User Acceptance Of Information Technology: Toward A Unified View
30. Benbasat I, Goldstein DK and Mead M (1987)	The Case Research Strategy In Studies Of Information-Systems
31. Shea S, DuMouchel W and Bahamonde L (1996)	A Meta-Analysis Of 16 Randomized Controlled Trials To Evaluate Computer-Based Clinical Reminder Systems For Preventive Care In The Ambulatory Setting
32. Goodhue DL and Thompson RL (1995)	Task-Technology Fit And Individual-Performance
33. Brancheau JC and Wetherbe JC (1987)	Key Issues In Information-Systems Management
34. Chomsky N (1956)	3 Models For The Description Of Language
35. Tajfel H (1974)	Social Identity And Intergroup Behaviour
36. Youla DC (1961)	On Factorization Of Rational Matrices
37. Klein HK and Myers MD (1999)	A Set Of Principles For Conducting And Evaluating Interpretive Field Studies In Information Systems
38. Orlikowski WJ (1993)	Case Tools As Organizational-Change – Investigating Incremental And Radical Changes In Systems-Development
39. Iacovou CL, Benbasat I and Dexter AS (1995)	Electronic Data Interchange And Small Organizations: Adoption And Impact Of Technology
40. Mata FJ, Fuerst WL and Barney JB (1995)	Information Technology And Sustained Competitive Advantage: A Resource-Based Analysis
41. Wrangham RW (1979)	Evolution Of Ape Social-Systems
42. Kane B and Sands DZ (1998)	Guidelines For The Clinical Use Of Electronic Mail With Patients
43. Bakos JY (1991)	A Strategic Analysis Of Electronic Marketplaces
44. Gefen D and Straub DW (1997)	Gender Differences In The Perception And Use Of E-Mail: An Extension To The Technology Acceptance Model

45. Schamber L, Eisenberg MB and Nilan MS (1990)	A Reexamination Of Relevance - Toward A Dynamic, Situational Definition
46. Thompson RL, Higgins CA and Howell JM (1991)	Personal Computing - Toward A Conceptual-Model Of Utilization
47. Spink A, Wolfram D, Jansen MBJ and Saracevic T (2001)	Searching The Web: The Public And Their Queries
48. White HD and Griffith BC (1981)	Author Cocitation - A Literature Measure Of Intellectual Structure
49. Bates MJ (1979)	Information Search Tactics
50. Belkin NJ (1980)	Anomalous States Of Knowledge As A Basis For Information-Retrieval
51. Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Rittenberg E, Burdick E, Seger DL, Vander Vliet M and Bates DW (1998)	Identifying Adverse Drug Events: Development Of A Computer-Based Monitor And Comparison With Chart Review And Stimulated Voluntary Report
52. Cimino JJ, Clayton PD, Hripcsak G and Johnson SB (1994)	Knowledge-Based Approaches To The Maintenance Of A Large Controlled Medical Terminology
53. Dickson GW, Leitheiser RL, Wetherbe JC and Nechis M (1984)	Key Information-Systems Issues For The 1980s
54. Bourdieu P (1975)	Specificity Of Scientific Field And Social Conditions Of Progress Of Reason
55. Huber GP (1984)	Issues In The Design Of Group Decision Support Systems
56. Orlikowski WJ (1996)	Improvising Organizational Transformation Over Time: A Situated Change Perspective
57. Mukhopadhyay T, Kekre S and Kalathur S (1995)	Business Value Of Information Technology - A Study Of Electronic Data Interchange
58. Karahanna E, Straub DW and Chervany NL (1999)	Information Technology Adoption Across Time: A Cross-Sectional Comparison Of Pre-Adoption And Post-Adoption Beliefs
59. Harter SP (1992)	Psychological Relevance And Information-Science
60. Saracevic T, Kantor P, Chamis AY and Trivison D (1988)	A Study Of Information Seeking And Retrieving .1. Background And Methodology
61. Webster J and Martocchio JJ (1992)	Microcomputer Playfulness - Development Of A Measure With Workplace Implications
62. Robertson SE	Probability Ranking Principle In IR
63. Saracevic T and Kantor P (1988)	A Study Of Information Seeking And Retrieving .3. Searchers, Searches, And Overlap
64. Ingwersen P (1996)	Cognitive Perspectives Of Information Retrieval Interaction: Elements Of A Cognitive Ir Theory
65. White HD and McCain KW (1998)	Visualizing A Discipline: An Author Co-Citation Analysis Of Information Science, 1972-1995
66. Srinivasan A (1985)	Alternative Measures Of System Effectiveness – Associations And Implications
67. Shannon CE (1956)	The Zero Error Capacity Of A Noisy Channel
68. Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM and Sheridan T (2001)	White Paper - Reducing The Frequency Of Errors In Medicine Using Information Technology
69. Bates MJ (1986)	Subject Access In Online Catalogs - A Design-Model

70. Vanrijsbergen CJ (1977)	Theoretical Basis For Use Of Co-Occurrence Data In Information-Retrieval
71. Compeau DR and Higgins CA (1995)	Application Of Social Cognitive Theory To Training For Computer Skills
72. Brancheau JC, Janz BD and Wetherbe JC (1996)	Key Issues In Information Systems Management: 1994-95 Sim Delphi Results
73. Schubert A, Glaänzel W and Braun T (1989)	Scientometric Datafiles - A Comprehensive Set Of Indicators On 2649 Journals And 96 Countries In All Major Science Fields And Subfields 1981-1985
74. Gallupe RB, Desanctis G and Dickson GW (1988)	Computer-Based Support For Group Problem-Finding - An Experimental Investigation
75. Barua A, Kriebel CH and Mukhopadhyay T (1995)	Information Technologies And Business Value - An Analytic And Empirical-Investigation
76. Agarwal R and Karahanna E (2000)	Time Flies When You're Having Fun: Cognitive Absorption And Beliefs About Information Technology Usage
77. Sanders GL and Courtney JF (1985)	A Field-Study Of Organizational-Factors Influencing Dss Success

Table App.2 presents, for each information scientist investigated in Chapter 5, the author's h-index in IS&LS and number of articles in IS&LS.

Table App.2: The h-indexes and number of publications in IS&LS of the first authors

Author	h	n	Author	h	n	Author	h	n
Salton G	18	48	Goodhue DL	8	14	Bakos JY	3	3
Saracevic T	17	106	Bourdieu P	7	12	Shannon CE	3	3
Schubert A	17	82	Orlikowski WJ	7	7	Tajfel H	3	3
Bates DW	17	62	Schamber L	6	20	Thompson RL	3	3
Benbasat I	17	40	Mukhopadhyay T	6	11	Huber GP	2	3
Bates MJ	17	34	Klein HK	6	10	Jha AK	2	3
Belkin NJ	16	36	Gefen D	6	9	Youla DC	2	3
Robertson SE	16	31	Barua A	6	8	Adams DA	2	2
Spink A	15	89	Brancheau JC	6	7	Chomsky N	2	2
Cimino JJ	14	76	Dickson GW	6	7	Hu M	2	2
Harter SP	14	74	Gallupe RB	6	7	Srinivasan A	2	2
White HD	14	36	Sanders GL	6	7	Taylor S	2	2
Ingwersen P	13	44	Jansen BJ	5	17	Webster J	2	2
Small H	13	35	Shea S	5	9	Iacovou CL	1	2
Straub DW	13	26	Karahanna E	4	7	Kane B	1	2
Vanrijsbergen CJ	11	25	Compeau DR	4	6	Wrangham RW	1	2
Dennis AR	11	19	Taylor RS	4	6	Daft RL	1	1
Egenhofer MJ	10	13	Price DJD	4	5	Gallager RG	1	1
Kuhlthau CC	9	23	Davis FD	4	4	Gruber TR	1	1
Doll WJ	9	17	Porter MF	4	4	Mata FJ	1	1
Haynes RB	8	18	Venkatesh V	3	7	Max J	1	1
Agarwal R	8	14	Deerwester S	3	5	Stockwell D	1	1

Table App.3 (referred to in Chapter 8) compares, for the most highly 36 articles in six subjects, the citations in the first six years after publication as a percentage of citations to date.

Table App.3: Citations in the first 6 years after publication as a percentage of all citations to date for the 36 most highly cited articles in each subject area

Rank by citations	Physics	Chemistry	Physiology	Economics	Statistics	Psychology
1	10.6	3.7	11.2	1.8	.3	10.9
2	.2	1.4	14.6	3.8	7.1	6.1
3	5.8	.7	13.9	5.2	7.7	3.7
4	9.8	12.1	16.0	3.8	3.6	29.1
5	9.2	8.0	18.4	5.4	2.0	3.7
6	5.3	31.3	35.6	4.3	5.5	22.7
7	3.0	25.1	6.2	3.7	9.1	20.7
8	1.5	16.1	27.0	6.3	3.2	5.8
9	1.9	17.3	5.5	9.3	16.5	7.5
10	9.9	2.9	26.6	16.5	3.9	11.9
11	.7	25.0	27.8	10.0	1.8	16.4
12	5.7	6.7	22.3	10.2	7.2	6.2
13	3.0	18.6	5.2	12.7	11.6	4.5
14	3.3	36.9	3.6	2.5	.7	17.8
15	5.3	5.7	11.9	3.8	10.3	10.1
16	13.1	25.0	35.3	16.8	20.1	8.5
17	5.5	19.4	11.7	11.3	1.7	29.8
18	19.8	18.1	4.6	15.7	3.1	6.1
19	16.1	.9	25.4	12.5	6.2	40.5
20	10.5	13.7	16.6	13.9	14.3	8.5
21	4.1	12.1	34.7	13.9	4.7	18.5
22	14.3	19.3	11.0	13.4	12.8	9.2
23	16.2	19.4	5.2	7.4	18.8	4.7
24	6.7	14.9	17.1	4.7	14.5	19.9
25	2.6	20.9	8.9	18.2	2.9	39.4
26	45.3	29.3	26.6	7.1	7.8	26.9
27	15.1	25.1	9.9	11.3	20.2	19.9
28	35.7	24.4	26.9	22.2	3.3	9.5
29	28.6	32.6	12.7	25.7	12.7	21.9
30	32.9	11.7	10.3	12.8	11.0	33.3
31	25.8	16.8	22.8	11.8	6.5	18.7
32	9.4	52.8	43.1	19.4	9.0	33.1
33	9.6	26.8	27.6	11.1	3.3	13.5
34	18.0	38.0	26.3	9.2	8.6	21.9
35	20.5	23.0	17.4	13.1	10.7	22.7
36	28.4	12.8	18.9	12.3	15.6	13.5
Mean	**12.6**	**18.6**	**18.3**	**10.6**	**8.3**	**16.6**

References

Ackerson, L.G. and Chapman, K. (2003). Identifying the role of multidisciplinary journals in scientific research. College & Research Libraries, 64(6), 468-478.

Adams, J. (2005). Early citation counts correlate with accumulated impact. Scientometrics, 63(3), 567-581.

Aksnes, D.W. (2003). Characteristics of highly cited papers. Research Evaluation, 12(3), 159–170.

Aleixandre, R., Valderrama, J., Desantes, J. and Torregrosa, A. (2004). Identification of information sources and citation patterns in the field of reciprocating internal combustion engines. Scientometrics, 59(3), 321-336.

Arencibia-Jorge, R., Barrios-Almaguer, I., Fernandez-Hernandez, S. and Carvajal-Espino, R. (2008). Applying successive H indices in the institutional evaluation: A case study. Journal of the American Society for Information Science and Technology, 59(1), 155-157.

Aversa, E.S. (1985). Citation patterns of highly cited papers and their relationship to literature aging - a study of the working literature. Scientometrics, 7(3-6), 383-389.

Avkiran, N.K. (1997). Scientific collaboration in finance does not lead to better quality research. Scientometrics, 39(2), 173-184.

Batista, P.D., Campiteli, M.G. and Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? Scientometrics, 68(1), 179-189.

Bordons, M., Zulueta, M.A., Romero, F. and Barrigon, S. (1999). Measuring interdisciplinary collaboration within a university: The effects of the Multidisciplinary Research Programme. Scientometrics, 46(3), 383-398.

Borgman, C.L. (1989). Bibliometrics and Scholarly Communication. Communication Research, 16(5), 583-599.

Borgman, C.L. ed. (1990). Scholarly Communication and Blbliometrics. Sage Publications, Newbury Park, California and London.

Borgman, C.L. and Furner, J. (2002). Scholarly communication and bibliometrics. Annual Review of Information Science and Technology, 36, 3-72.

Borgman, C.L. and Rice, R.E. (1992). The convergence of information-science and communication - a bibliometric analysis. Journal of the American Society for Information Science, 43(6), 397-411.

Bornmann, l. and Daniel, H.D. (2007). What do we know about the h index? Journal of the American Society for Information Science and Technology, 58(9), 1381-1385.

Bradford, S.C. (1934). Sources of information on specific subjects. British Journal of Engineering, 137(1), 85-86

Braun, T., Glänzel, W. and Schubert, A. (2006). A Hirsch-type index for journals. Scientometrics, 69(1), 169-173.

Burrell, Q.L. (2005). Are "sleeping beauties" to be expected? Scientometrics, 65(3), 381-389.

Cano, V. and Lind, N.C. (1991). Citation life-cycles of 10 citation-classics. Scientometrics, 22(2), 297-312.

Carlin, A. (2003). Some Bibliographic Practices in Interdisciplinary Work: Accounting for Citations in Library and Information Sciences. Accountability in Research: Policies & Quality Assurance, 10(1), 27-45.

Chan, H.C., Kim, H.W. and Tan, W.C. (2006). Information systems citation patterns from international conference on information systems articles. Journal of the American Society for Information Science and Technology, 57(9), 1263-1274

Clausen, H. and Wormell, I. (2001) A bibliometric analysis of IOLIM conferences 1977-1999. Journal of Information Science, 27(3), 157-169.

Crase, D. and Rosato, F.D. (1992). Single versus multiple authorship in professional journals. Journal Of Physical Education, Recreation And Dancing, 63(7), 28-31.

Cronin, B. (1984). The citation process. Taylor Graham, London.

Cronin, B. and Atkins, H.B. ed. (2000). The Web of Knowledge: A Festschrift in Honor of Eugene Garfield. ASIS Monograph Series, Information Today, Medford, New Jersey.

Cronin, B. and Meho, L. (2007). Timelines of Creativity: A Study of Intellectual Innovators in Information Science. Journal of the American Society for Information Science and Technology, 58(13), 1-12.

Cronin, B. and Meho, L. (2006). Using the h-index to rank influential information scientists. Journal of the American Society for Information Science and Technology, 57(9), 1275–1278.

Cronin, B., Shaw, D. and La Barre, K. (2003). A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. Journal of the American Society for Information Science and Technology, 54(9), 855-871.

Crow, G.M., Levine, L. and Nager, N. (1992). Are 3 heads better than one - reflections on doing collaborative interdisciplinary research. American Educational Research Journal, 29(4), 737-753.

Csajbok, E., Berhidi, A., Vasas, L. and Schubert, A. (2007). Hirsch-index for countries based on essential science indicators data. Scientometrics, 73(1), 91-117.

Egghe, L. (2006). Theory and practise of the g-index. Scientometrics, 69(1), 131-152.

Eto, H. (2003). Interdisciplinary information input and output of a nano-technology project. Scientometrics, 58(1), 5-33.

Fox, M.F. and Faver, C.A. (1984). Independence and cooperation in research - the motivations and costs of collaboration. Journal Of Higher Education, 55(3), 347-359.

Garfield, E. (1993). From Obliteration to Immortality. Current Contents, 45, 391-92.

Garfield, E. (1985a). The articles most cited in the SCI from 1961 to 1982 .7. Another 100 citation-classics - the Watson-Crick double helix has its turn. Current Contents, 20, 3-12.

Garfield, E. (1985b). The articles most cited in the SCI from 1961 to 1982 .8. Ninety-eight more classic papers from unimolecular reaction velocities to natural opiates - the changing frontiers of science. Current Contents, 33, 3-11.

Garfield, E. (1980). Premature discovery or delayed recognition - Why? Current Contents, 26, 5-10.

Garfield, E. (1979). Citation Indexing: Its Theory and Application in Science, Technology, and Humanities. John Wiley & Sons, New York.

Garfield, E. (1975). The 'obliteration phenomenon' in science—and the advantage of being obliterated! Current Contents, 51/52, 5-7.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. and Trow, M. (1994). The new production of knowledge: the dynamics of science and research in contemporary societies. London: Sage.

Glänzel, W. (2002). Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. Library Trends, 50(3), 461-473.

Glänzel, W. (2000). Science in Scandinavia: A bibliometric approach. Scientometrics, 48(2), 121-150.

Glänzel, W. and Garfield, E. (2005). The myth of delayed recognition. The Scientist, 18(11), 8.

Glänzel, W. and Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. Scientometrics, 56(3), 357-367.

Glänzel, W., Schlemmer, B. and Thijs, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. Scientometrics, 58(3), 571-586.

Glänzel, W., Schubert, A. and Czerwon, H.J. (1999). An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. Scientometrics, 44(3), 427-439.

Gómez, I., Bordons, M., Fernandez, M.T. and Mendez, A. (1996). Coping with the problem of subject classification diversity. Scientometrics, 35(2), 223-235.

Gómez, I., Fernandez, M.T. and Sebastian, J. (1999). Analysis of the structure of international scientific cooperation networks through bibliometric indicators. Scientometrics, 44(3), 441-457.

Hart, R.L. (2007). Collaboration and article quality in the literature of academic librarianship. Journal Of Academic Librarianship, 33(2), 190-195.

Haythornthwaite, C. (2006). Learning and knowledge networks in interdisciplinary collaborations. Journal of the American Society for Information Science and Technology, 57(8), 1079-1092.

Herbertz, H. (1995). Does it pay to cooperate - a bibliometric case-study in Molecular-Biology. Scientometrics 33(1), 117 1995.

Herring, S.D. (1999). The value of interdisciplinarity: A study based on the design of Internet search engines. Journal of the American Society for Information Science, 50(4), 358-365.

Hinze, S. (1999). Collaboration and cross-disciplinarity in autoimmune diseases. Scientometrics, 46(3), 457-471.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46), 16569-16572.

Iglesias, J.E. and Pecharroman, C. (2007). Scaling the h-index for different scientific ISI fields. Scientometrics, 73(3), 303-320.

Katz, J.S. and Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. Scientometrics, 40(3), 541-554.

Katz, J.S. and Martin, B.R. (1997). What is research collaboration? Research Policy, 26(1), 1-18.

Le Pair, C. (1980). Switching between academic disciplines in universities in the Netherlands. Scientometrics, 2(3), 177-191.

Leimu, R, (2005). Does scientific collaboration increase the impact of ecological articles? Bioscience 55 (438).

Levitt, J.M. and Thelwall, M. (2008a). Indicators for the effectiveness of the Web of Science subject categories: A case study of Library and Information Science. An oral presentation at the 10th International Science and Technology conference in Vienna.

Levitt, J.M. and Thelwall, M. (2008b). The impact of late citation on evaluating applied research: A case study of the Scopus 'Engineering' category. A poster at the 10th International Science and Technology conference in Vienna.

Levitt, J.M. and Thelwall, M. (2007a). Atypical Citation Patterns in the Twenty Most Highly Cited Documents in Library and Information Science. Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics: Madrid (Spain).

Levitt, J.M. and Thelwall, M. (2007b). Two new indicators derived from the h-index for comparing citation impact: Hirsch frequencies and the Normalised Hirsch Index. Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics: Madrid (Spain).

Levitt, J.M. and Thelwall, M. (in press 2008a). The most highly cited Library and Information Science articles: Interdisciplinarity, first authors and citation patterns. This article is scheduled to be publication in Scientometrics, 78 (1). Available online at http://www.scit.wlv.ac.uk/~cm1993/highly_cited_LIS.doc (last visited September 5, 2008).

Levitt, J.M. and Thelwall, M. (in press 2008b). Patterns of annual citation of highly cited articles and the prediction of their citation ranking: A comparison across subjects. This article is scheduled to be publication in Scientometrics, 77(1).

Levitt, J.M. and Thelwall, M. (in press 2008c). Is multi-disciplinary research more highly cited? A macro-level study. In May 2008 this article was accepted for publication in JASIsT.

Levitt, J.M. and Thelwall, M. (under review). Citation levels and collaboration within Library and Information Science. In July 2008 this article was submitted to JASIST.

Lewison, G. and Paraje, G. (2004). The classification of biomedical journals by research level. Scientometrics, 60(2), 145-157.

Leydesdorff, L. (2008). Caveats for the use of citation indicators in research and journal evaluations. Journal of the American Society for Information Science and Technology, 59(2), 278-287.

Leydesdorff, L. (2007). Mapping interdisciplinarity at the interfaces between the science citation index and the social science citation index. Scientometrics, 71(3), 391-405.

Leydesdorff, L. (2004). Top-down decomposition of the Journal Citation Report of the Social Science Citation Index: Graph- and factor-analytical approaches. Scientometrics, 60(2), 159-180.

Leydesdorff, L. and Etzkowitz, H. (2001). A triple helix of University-Industry-Government relations: Mode 2 and the globalization of national systems of innovation, in: Science under pressure. Aarhus: The Danish Institute for Studies in Research and Research Policy. Available online at http://www.afsk.au.dk/ftp/Science_under_pressure/2001_1.pdf (last visited September 5, 2008).

Martin, B.R. (1996). The use of multiple indicators in the assessment of basic research. Scientometrics, 36(3), 343-362.

Mccain, K.W. (2007). Analysing Influence Over Time: An Historiographic Mapping of the Research of Conrad Hal Waddington (1905-1975). Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics: Madrid (Spain).

Mccain, K.W. and Turner, K. (1989). Citation context analysis and aging patterns of journal articles in Molecular-Genetics. Scientometrics, 17(1-2), 127-163.

Meho, L.I. and Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus Scopus and Google Scholar. Journal of the American Society for Information Science and Technology, 58(13), 2105-2125.

Merton, R K. (1968). On the shoulders of giants: A Shandean postscript. New York: Harcourt Brace & World.

Meyer, T. and Spencer, J. (1996). A citation analysis study of library science: Who cites librarians? College and Research Libraries, 57(1), 23-33.

Minasny, B., Hartemink, A.E. and McBratney, A. (2007). Soil science and the h index. Scientometrics, 73(3), 257-264.

Moed, H.F. (2005). Citation analysis in research evaluation. Kluwer Academic Publishers, New York.

Moed, H.F., Van Leeuwen, T.N. and Reedijk, J. (1999). Towards appropriate indicators of journal impact. Scientometrics, 46(3), 575-589.

Morillo, F., Bordons, M. and Gómez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. Journal of the American Society for Information Science and Technology, 54(13), 1237-1249.

Morillo, F., Bordons, M. and Gómez, I. (2001). An approach to interdisciplinarity bibliometric indicators. Scientometrics, 51(1), 203-222.

Norris, M. and Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. Journal of Informetrics, 1(2), 161-169.

Oppenheim, C. (2007). Using the h-index to rank influential British researchers in information science and librarianship. Journal of the American Society for Information Science and Technology, 58(2), 297-301.

Pereira, J.C.R., Fischer, A.L. and Escuder, M.M.L. (2000). Driving factors of high performance in Brazilian Management Sciences for the 1981-1995 period. Scientometrics, 49(2), 307-319.

Persson, O., Glänzel, W. and Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. Scientometrics, 60(3), 421-432 2004.

Pierce, S.J. (1999). Boundary crossing in research literatures as a means of interdisciplinary information transfer. Journal of the American Society for Information Science, 50(3), 271-279.

Porter, A.L. and Chubin, D.E. (1985). An indicator of cross-disciplinary research. Scientometrics, 8(3-4), 161-176.

Porter, A.L., Cohen, A.S., Roessner, J.D. and Perreault, M. (2007). Measuring researcher interdisciplinarity. Scientometrics, 72(1), 117-147.

Price, D. J. (1963). Little Science, Big Science. Columbia University Press, New York.

Pudovkin, A.I. and Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. Journal of the American Society for Information Science and Technology, 53(13), 1113-1119.

Qin, J., Lancaster, F.W. and Allen, B. (1997). Types and levels of collaboration in interdisciplinary research in the sciences. Journal of the American Society for Information Science, 48(10), 893-916.

Rafols, I. and Meyer, M. (2007). How cross-disciplinary is bionanotechnology? Explorations in the specialty of molecular motors. Scientometrics, 70(3), 633-650.

Rice, R.E. and Crawford, G.A. (1992). Analysis of citations between communication and library and information science articles. Proceedings of the 55th Annual Meeting of the American Society for Information Science, 1992, 8-12.

Rinia, E.J., Van Leeuwen, T.N. and Van Raan, A.F.J. (2002). Impact measures of interdisciplinary research in physics. Scientometrics, 53(2), 241-248.

Rinia, E.J., Van Leeuwen, T.N., Bruins, E.E.W., Van Vuren, H.G. and Van Raan, A.F.J. (2002). Measuring knowledge transfer between fields of science. Scientometrics, 54(3), 347-362.

Rinia, E.J., Van Leeuwen, T.N., Bruins, E.E.W., Van Vuren, H.G. and Van Raan, A.F.J. (2001). Citation delay in interdisciplinary knowledge exchange. Scientometrics, 51(1), 293-309.

Rousseau, R. and Zuccala, A. (2004). A classification of author co-citations: Definitions and search strategies. Journal of the American Society for Information Science and Technology, 55(6), 513-529.

Schubert, A. (2007). Successive h-indices. Scientometrics, 70(1), 201-205.

Shama, G., Hellgardt, K. and Oppenheim, C. (2000). Citation footprint analysis Part I: UK and US chemical engineering academics. Scientometrics, 49(2), 289-305.

Steele, T.W. and Stier, J.C. (2000). The impact of interdisciplinary research in the environmental sciences: A forestry case study. Journal of the American Society for Information Science, 51(5), 476-484.

Stock, M. and Stock, W.G. (2006). Intellectual property information: A comparative analysis of main information providers. Journal of the American Society for Information Science and Technology, 57(13), 1794-1803.

Tang, R. (2004a). Evolution of the interdisciplinary characteristics of information and library science. Proceedings of the 67th ASIST Annual Meeting, 41(1), 54-63.

Tang, R. (2004b). Visualizing interdisciplinary citations to and from information and library science publications. Eighth International Conference on Information Visualisation, Proceedings, 972-977.

Tenopir, C. (2001). Online Databases-The Power of Citation Searching. Library Journal, 126(18), 39-40.

Uthman, O.A. (2008). HIV/AIDS in Nigeria: A bibliometric analysis. BMC Infectious Diseases, 8 (3).

Van Dalen, H.P. and Henkens, K. (2005). Signals in science - On the importance of signaling in gaining attention in science. Scientometrics, 64(2), 209-233.

Van Leeuwen, T.N. and Tijssen, R. (2000). Interdisciplinary dynamics of modern science: analysis of cross-disciplinary citation flows. Research Evaluation, 9(3), 183-187.

Van Leeuwen, T.N., Van der Wurff, L.J. and Van Raan, A.F.J. (2001). The use of combined bibliometric methods in research funding policy. Research Evaluation, 10(3), 195-201.

Van Raan, A.F.J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics, 67(3), 491-502.

Van Raan, A.F.J. (2004). Sleeping beauties in science. Scientometrics, 59(3), 467-472.

Van Raan, A.F.J. (1998). The influence of international collaboration on the impact of research results - Some simple mathematical considerations concerning the role of self-citations. Scientometrics, 42(3), 423-428.

Walters, W.H. (2007). Google Scholar coverage of a multidisciplinary field. Information Processing & Management, 43(4), 1121-1132.

Wang, Y., Wu, Y.S., Pan, Y.T., Ma, Z. and Rousseau, R. (2005). Scientific collaboration in China as reflected in co-authorship. Scientometrics, 62(2), 183-198.

White, H.D. (2001). Authors as citers over time. Journal of the American Society for Information Science and Technology, 52(2), 87-108.

White, H.D. and McCain, K.W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. Journal of the American Society for Information Science, 49(4), 327-355.

Yi, H., Ao, X.L. and Ho, Y.S. (2008). Use of citation per publication as an indicator to evaluate pentachlorophenol research. Scientometrics, 75(1), 67-80.

Zitt, M., Ramanana-Rahary, S. and Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. Scientometrics, 63(2), 373-401.