

# Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis

**Kayvan Kousha**

*Department of Library and Information Science, University of Tehran, Iran and School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1ST, United Kingdom. E-mail: kkoosha@ut.ac.ir*

**Mike Thelwall**

*School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1ST, United Kingdom. E-mail: m.thelwall@wlv.ac.uk*

**We use a new data gathering method, “Web/URL citation,” Web/URL and Google Scholar to compare traditional and Web-based citation patterns across multiple disciplines (biology, chemistry, physics, computing, sociology, economics, psychology, and education) based upon a sample of 1,650 articles from 108 open access (OA) journals published in 2001. A Web/URL citation of an online journal article is a Web mention of its title, URL, or both. For each discipline, except psychology, we found significant correlations between Thomson Scientific (formerly Thomson ISI, here: ISI) citations and both Google Scholar and Google Web/URL citations. Google Scholar citations correlated more highly with ISI citations than did Google Web/URL citations, indicating that the Web/URL method measures a broader type of citation phenomenon. Google Scholar citations were more numerous than ISI citations in computer science and the four social science disciplines, suggesting that Google Scholar is more comprehensive for social sciences and perhaps also when conference articles are valued and published online. We also found large disciplinary differences in the percentage overlap between ISI and Google Scholar citation sources. Finally, although we found many significant trends, there were also numerous exceptions, suggesting that replacing traditional citation sources with the Web or Google Scholar for research impact calculations would be problematic.**

## Introduction

The partial transition of academic publishing from print to the Web has been a key factor in motivating information professionals to explore online communication patterns (e.g., Fry, 2004; Kling & McKim, 1999). In particular, many have assessed whether the methods of bibliometrics, such as

citation analysis, can be applied to the Web environment (e.g., Almind & Ingwersen, 1997; Borgman & Furner, 2002; Ingwersen, 1998; Rousseau, 1997). While early studies tended to analyze links to journal Web sites or online articles (Harter & Ford, 2000; Smith, 1999; Vaughan & Hysen, 2002; Vaughan & Thelwall, 2003), later research switched to prefer text-based citations in Web pages (Kousha & Thelwall, in press; Vaughan & Shaw, 2005; Vaughan & Shaw, 2003).

From the early 1990s, researchers have discussed the potential for open access (OA) publishing (e.g., free access online journals) to revolutionize scholarly communication (e.g., Harnad, 1990; Harnad, 1991; Harter, 1996; Harnad, S., 1999) and explored author experiences and opinions about publishing in open access journals and self-archiving (e.g., Swan & Brown, 2004; Swan & Brown, 2005). The next natural step was to seek evidence for the impact of OA publishing using existing bibliometric techniques (as described in Borgman & Furner, 2002) and in this regard researchers have shown that the online availability of articles associates with higher citation counts in several subject areas (Antelman, 2004; Harnad & Brody, 2004; Kurtz, 2004; Lawrence, 2001; Shin, 2003). The increasing number of OA journals indexed in the Thomson Scientific (formerly Thomson ISI, and for convenience referred to as ISI here) citation databases (more than 200 at the time of this study) not only supports their acceptance as a valid outlet for publishing scientific articles but also allows researchers to use ISI citations as a measure of assessment (Brody et al., 2004) or to compare OA and non-OA journals impact across many disciplines (ISI press release, 2004).

The ISI has for a long time managed the preeminent international, multidisciplinary database for citation tracking. Nevertheless, the significant degree of open access publishing in fields such as computer science and physics has allowed some Web-based repositories to be used as an alternative for assessing the citation impact of articles (see literature

Received May 8, 2006; revised July 24, 2006; accepted July 24, 2006

© 2007 Wiley Periodicals, Inc. • Published online 29 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20584

review below). In addition, researchers have also developed novel hyperlink-based methods for impact assessment based upon the “whole Web,” leveraging analogies with citations and using commercial search engines for extracting link data (see Thelwall, Vaughan & Björneborn, 2005). Hence, search engine indexes can be used, in practice, as alternative citation databases. Both links and citations are inter-document connections, and high numbers of inlinks (Brin & Page, 1998) and citations (Moed, 2005) are both being regarded as positive indicators of value. Commercial search engines have also been used to extract Web citations, e.g., mentions of journal articles titles in Web pages (Vaughan & Shaw, 2003; Vaughan & Shaw, 2005), and URL citations, which are counts of the number of times the URL of a resource is mentioned in other Web pages (Kousha & Thelwall, 2006). As with the above-mentioned repository research, these studies compared their results with ISI citations as a scholarly source with better-known value and validity. In most cases, the Web-based citations correlated with ISI citations but with significant differences in the total numbers of citations found and with some exceptions. On the basis of these findings, there have been claims that the Web could be an alternative to the ISI for citation impact calculations (Vaughan & Shaw, 2005). Nevertheless, there are differences in the extent to which disciplines publish on the Web and write journal articles (Fry & Talja, 2004; Kling & McKim, 1999), and so more information is needed about disciplinary differences in online citation counting to confirm or deny these claims. This may also shed light on the strengths and weaknesses of the ISI’s coverage of the scholarly literature.

In the present study, we explore the commonality between conventional and Web-extracted citation patterns for open access journals in some science and social science disciplines, incorporating both Google Scholar (<http://scholar.google.com>) and a new Web/URL citation method. Hence, we identify and analyze disciplinary differences within and between traditional and Web-based citation counts on a broader level than has previously been attempted.

## Related Studies

There is now a considerable body of quantitative research into scholarly use of the Web, as reviewed in the recent *Annual Review of Information Science and Technology* (ARIST) Webometrics chapter (Thelwall, Vaughan & Björneborn, 2005). Link analysis is particularly developed field, but some research has also used Web/URL citations.

### Link Analysis

Most information science *link analysis* studies have been motivated by citation analysis, for example, exploring analogies between citations and Web links (Smith, 2004), using the term “situation” to refer to a cited Web site (Rousseau, 1997) and defining the “Web Impact Factor” as a Web counterpart of the ISI’s Impact Factor for journals (Ingwersen, 1998). While some information scientists have emphasized the structural

similarity between linking and citing (Borgman & Furner, 2002), others have instead highlighted their differences (e.g., Björneborn & Ingwersen, 2001; Egghe, 2000; Glanzel, 2003). This debate is not yet closed.

Correlation tests have been used as an indirect approach to assess the extent of the agreement between traditional and Web-based citation patterns. Correlation tests typically take the form of comparing two sets of numbers, such as Web and ISI citations to the same collection of journal articles, revealing the extent to which larger values from one source associate with larger values from the other source. A high degree of correlation could indicate that one causes or influences the other (e.g., if ISI citations sometimes appear because scholars found references online) or that the two have a common underlying influence (e.g., if both tend to reflect the academic impact of the cited work). This indirect approach is useful as a kind of shortcut to understanding what Web measurements may represent by comparing them with better-known statistics. As with citation analysis, direct approaches, such as a content analysis and interviewing Web authors, are also needed for the effective interpretation of Web-based variables, however (Oppenheim, 2000; Thelwall, 2006).

Smith (1999) was one of the first researchers to use link analysis techniques to examine the relationship between inlinks and ISI Impact Factors, finding no significant association for 22 Australasian refereed e-journals. Similarly, Harter and Ford (2000) compared links to 39 scholarly e-journals with ISI citations and found no significant correlation between link counts and ISI impact factors. Although most studies applied quantitative methods (mainly correlation tests), Kim (2000) and Herring (2002) applied qualitative methods to explore motivations for creating links in e-journal articles, finding both overlaps with traditional citer motivations and some new electronic medium-specific reasons. The first study to produce a statistically significant result was that of Vaughan and Hysen (2002), finding a correlation between the number of links to a journal Web site and the associated journal Impact Factor for ISI-indexed library and information science (LIS) journal Web sites. Perhaps this research was successful because it was discipline-specific, even though it was dominated by non-OA journals. It was also able to take advantage of the fact that by the time of the study, most mainstream journals seemed to have deployed an associated Web site, which was probably not true at the time of the early OA studies. Follow-up research confirmed the correlation and showed that journals with more online content tended to attract more links, as did older journal Web sites in both law and library and information science (Vaughan & Thelwall, 2003).

## Web Citations

In the above-mentioned studies, Web links were the online variable, but Vaughan and Shaw (2003) subsequently used *Web citations* as impact assessment measures for journals. They compared ISI citations to library and information science journal articles with citations in the Web, using search engine

searches to count the number of times each selected journal article title was mentioned in Web pages (i.e., not necessarily a full bibliographic citation with author names, journal name, etc.). They found significant correlations, suggesting that online and offline citation impacts could be in some way similar phenomena, and hinting that the Web via search engines could be a possible replacement for the ISI citation databases. In a follow-up study, they found relationships between ISI and Web citations to articles from 114 biology, genetics, medicine, and multidisciplinary science journals, confirming that their earlier results were widely applicable to the hard sciences. They also classified Web citations using a predefined scheme to examine the proportion of Web citations reflecting the intellectual impact of the articles (Vaughan & Shaw, 2005). Most of their selected journals were ISI journals with independent Web sites that were not open access. They concluded that Web and ISI citation counts measured a similar level of impact.

As mentioned in the Introduction section, there are relatively few comprehensive studies across several subject areas comparing conventional citations (e.g., ISI citations) with Web-based citations at the article or journal level, with the latter Vaughan and Shaw (2005) study being an exception. In order to tackle the specific issue of disciplinary differences, for example, Van Impe and Rousseau (2006) conducted a similar comparison for some Dutch and French humanities journals but found very few Web or ISI citations to these and hence were unable to draw strong conclusions.

### *Specialist Digital Libraries*

An opportunity to study alternative sources to the ISI for citations is afforded by the current crop of digital libraries. CiteSeer (<http://citeseer.ist.psu.edu/>), for instance, is an index of primarily computing journal and conference articles culled from the Web. It also generates formal citations from the bibliographic references in the online scholarly articles that it indexes. Goodrum, McCain, Lawrence, and Giles (1999) used this data to compare citation patterns in online computer science articles indexed in CiteSeer with citations from the ISI. One significant difference was that in computer science, the citations of conference articles seem to be under-represented by the ISI (Goodrum, et al., 2001). It is not clear whether this is desirable, however, given that the ISI applies quality control mechanisms to select journals for inclusion in their databases, something that does not apply to the Web as a whole. Zhao and Logan (2002) conducted a similar study of the XML research area and found that CiteSeer provided more citations than the ISI for this relatively new and fast moving field, a pointer to a possible source of disciplinary differences in online citation patterns. A later study found a less than 10% overlap between ISI and CiteSeer citations for XML research (Zhao, 2005). Other researchers have investigated citation behavior in a variety of other digital libraries, including comparisons with usage statistics to see whether highly cited articles are also highly read (Harnard & Carr, 2000; Kurtz et al., 2005).

## **Google Scholar**

The citation facility of Google Scholar is a potential new tool for bibliometrics. Launched in November 2004, Google Scholar claims to include “peer-reviewed articles, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations” (About Google Scholar, 2005). Perhaps some of these documents would not otherwise be indexed by search engines such as Google, so they would be “invisible” to Web searchers, and clearly some would be similarly invisible to *Web of Science* users because it is dominated by academic journals.

Jasco (2004, 2005a, 2005b) has noticed both uneven coverage of scholarly publishers’ archives and false matches reported by the early Google Scholar. Nevertheless, its use has been claimed to be “commonplace amongst all sectors of the academic community” because of “ease of use, saving time, and access to a wide range of resources” (Friend, 2006). The following publishers have heralded it because of its coverage of academic information: ACM, Annual Reviews, arXiv, Blackwell, IEEE, Ingenta, the Institute of Physics, NASA Astrophysics Data System, PubMed, Nature Publishing Group, RePEc (Research Articles in Economics), Springer, and Wiley Interscience (Notess, 2005). Many Web sites from universities and nonprofit organizations are also included—most notably, the OCLC Open WorldCat, with millions of bibliographic records (Notess, 2005). Can researchers and students, then, especially those who have no access to conventional fee-based citation indexes, such as the *Web of Science* and *Scopus*, use Google Scholar for locating scholarly information? Previous research has suggested that 72% of authors used Google to search the Web for scholarly articles (Swan & Brown, 2005); hence, it can be expected that a considerable number of researchers and students would be willing to try Google Scholar.

Only a few reported studies have compared the ISI and Google Scholar for citation impact calculations. Bauer and Bakkalbasi (2005) compared the citation counts provided by the ISI *Web of Science*, Elsevier’s *Scopus* abstract and indexing database, and Google Scholar for articles from the Journal of the American Society for Information Science and Technology (JASIST) published in 1985 and 2000. For articles published in 2000, Google Scholar provided significantly higher citation counts than either the *Web of Science* or *Scopus*, while there was no significant difference between the *Web of Science* and *Scopus*. The authors didn’t apply statistical tests for the year 1985, however, because of the high number of missing records. Pauly and Stergiou (2005) compared citations from the ISI and Google Scholar to 99 articles in 11 disciplines as well as 15 highly cited articles. Each discipline was represented by three authors, and each author was represented by three high-, medium-, and low-cited articles. The results suggested that the ISI and Google Scholar results were approximately equal for articles published after 1990, but these findings are suggestive rather than conclusive due to the small number of authors represented. Belew (2005)



selected six academics at random and compared citations to publications by these authors indexed by the ISI with those reported by Google Scholar. Again, the small number of academics prevents generalization, but it is noteworthy that only a small minority of citations found were in both sources, i.e., there was a small overlap.

So far, nobody seems to have investigated citation matches from a disciplinary differences perspective. Perhaps systematic studies of Google Scholar are difficult because of “minimal information about the content of Google Scholar,” such as publisher lists, journal lists, time span, or the disciplinary distribution of records (Jacso, 2005b). For this reason the extent to which these findings are generalizable is not known and more comparisons between ISI and Google Scholar citation impact are needed.

### *URL Citations*

*URL citation analysis*, counting the number of Web pages containing the URL of an OA article, is different from both link analysis and Web citation analysis, although an URL citation is also a link when the visible URL is also a hyperlink. One advantage of URL citation analysis over Web citation analysis is that multiple articles may have the same title but not the same URL. A disadvantage is that both links and citations may exclude a visible URL citation, and so URL citations capture only a proportion of times an article is referred to online (although MSN Search can simultaneously capture hyperlinks and URL citations; Stuart, personal communication, April 2006). In fact, URL citations, links, and Web citations all overlap to some extent but not completely. A previous comparison of peer-reviewed open access LIS journal articles found a significant correlation between ISI and URL citation counts and also between average numbers of ISI and URL citations (Kousha & Thelwall, in press). A classification of URL citations in that study estimated that 43% reflected citation-like intellectual impact and a further 18% represented informal scholarly uses. Hence, URL citations are at first blush a reasonable source of citation counts.

### **Research Questions**

Although many of the above-reviewed studies used either online citation indexes or search engine data, no comprehensive research so far has examined refereed open access journals across the sciences and social sciences to seek disciplinary differences in the relationships between ISI and different types of Web/URL citations (defined more precisely below in the methods section) or Google Scholar.

We address the three questions below to compare ISI and different types of Web-extracted citation patterns at the individual article and journal level. We focus on just OA journals to allow the widest possible type of Web citations (i.e., including URL citations) and to give an even basis for comparison. We use correlation tests as an indirect approach for interpreting Web-extracted citation counts and direct interpretation to validate the quantitative results.

1. Do Google Scholar citations to journal articles correlate with ISI citations in all disciplines?
2. Do Web-based citations (more precisely, Web/URL citations, defined below) to journal articles correlate with ISI citations in all disciplines?
3. Are there clear disciplinary differences between conventional and Web-based citation patterns?

### **Methods**

#### *Discipline, Journal, and Article Selection*

For the purpose of this study, OA journals are restricted to freely accessible English language journals available on the Web (in electronic only or both in electronic and print formats) with articles that have undergone some kind of peer or editorial review process—irrespective of whether the electronic publishing is the primary or secondary medium for the journal (see Kling & Callahan, 2003). We needed OA journals that had been published at least since 2001 in order to allow a significant time window in which to attract citations. An initial study based upon the Directory of Open Access Journals ([www.doaj.org](http://www.doaj.org)) and an ISI essay on the impact of open access journals (ISI press release, 2004) showed that there were few science and social science disciplines having enough satisfactory open access journals published in 2001. This low rate of OA journal publication in many areas is a limitation of our study and is mentioned again in the discussion. We used both of the above-mentioned sources and other related directories to locate OA journals. Ulrich’s Periodical Directory (2004) was consulted for official journal Web site URLs and the availability of OA journals in electronic only or in both electronic and print formats. The study used only the official Web sites of OA journals (the journal publisher’s Web site) for recording article URLs. Therefore, URLs of articles in mirror sites were not examined. OA journals that didn’t have an independent Web site were also included in this study, because the data collection method (Google searches) located Web/URL citations to journal articles in the text of other Web pages, rather than links to whole journal Web sites. If a journal Web site was in an individual HTML “frame” and all its articles had the same URL, then we excluded it from this study. Our final sample included 108 open access journals, 55 of which were indexed in the ISI *Web of Science* at the time of this study (Table 1).

The factors that were considered when selecting the science and social science disciplines included the number of refereed or editor-reviewed OA journals in each discipline and the accessibility of journal Web sites for data gathering. We selected biology, chemistry, physics, and computer science to represent a range of (hard) *sciences*, and we selected economics, education, sociology, and psychology to represent a range of *social sciences*. The selection of four science and four social science disciplines allowed comparisons between broadly similar disciplines as well as between distinctly different ones.

For each selected journal, we applied a sampling method for a systematic selection of journal articles (omitting

TABLE 1. Statistics for the sample collection of journal articles.

Discipline	Selected journals	Sampled research articles	ISI-indexed journals in sample	Sampled journals also available in print form
Biology	21	325 (19.7%)	17	13 (62%)
Chemistry	15	325 (19.7%)	15	12 (80%)
Physics	16	325 (19.7%)	12	11 (69%)
Computing	12	183 (11.1%)	5	9 (75%)

reports, editorials, and book reviews). The exact title of each research article was recorded, along with its (HTML or PDF) URL. For a statistically representative selection, in each discipline we took a random sample proportional to the total number of articles in each journal. Consequently, in each discipline, journals with more published articles had more articles in our final sample of 1650 research articles.

Using *Ulrich's Periodical Directory* and information in journal Web sites, we found that 52% (56 of 108) of the selected journals were available in both print and electronic formats and 48% (52) were exclusively available online (Table 1). See [http://cybermetrics.wlv.ac.uk/articledata/Kousha\\_1.doc](http://cybermetrics.wlv.ac.uk/articledata/Kousha_1.doc) for the journal names.

#### *ISI, Google Scholar, and Google Web/URL Citation Counts*

For the 55 ISI-indexed journals (at the time of this study), the number of citations to each article in the sample was recorded, as reported by the *Web of Science*. Since many open access journals weren't indexed (53 titles) or were indexed after 2001, their names were searched for in the "Cited Reference Search" field of the *ISI Web of Science* as an alternative way to find out the number of citations their 2001 articles had received from other ISI-indexed journal articles. In order to prevent possible errors due to similar abbreviations for different journals, the first author name and volume of each retrieved article was checked against the original OA article in the sample. This method is similar to that applied by Vaughan and Shaw (2005) for ISI-indexed journals and latter by Kousha and Thelwall (2006) for journals not indexed by the ISI. For the purpose of this study, both searches were limited to citations to year 2001 articles and journal names were truncated if necessary.

For Google Scholar citation counts, we searched the titles (taken from journal Web site tables of contents) of all 1,650 sampled articles as phrase searches in the main Google Scholar search page. We found that some titles with mathematical or chemical formulae did not retrieve any matches if their complete titles were used. Thus, it was necessary to omit a portion of some article titles (especially in physics, chemistry, and biology) during the search process in order to generate effective searches. We manually checked the search results against the original citation information to avoid false matches and to remove any duplicate citing documents. We then recorded the number of Google Scholar citations by clicking the "cited by" option below each retrieved record after omitting incorrect matches.

Google was chosen for extracting Web/URL citation counts because the results of previous studies have shown that it provided the most comprehensive (Bar-Ilan, 2004) and stable search results over time (Vaughan, 2004; Vaughan & Shaw, 2005). Google had good coverage of HTML and non-HTML documents and supported the syntax necessary for extracting both Web and URL citations at the time of this study, as described below. Nevertheless, the results of Google do not represent the whole Web (Lawrence & Giles, 1999), only the portion of the Web that it has crawled and reports for user searches (Bar-Ilan, 1999; Mettrop & Nieuwenhuysen, 2001).

For what we call Google *Web/URL citations*, methods from two previous studies were used. The article title phrase search method of Vaughan and Shaw (2005) for retrieving *Web citations* and the URL search method used by Kousha and Thelwall (2006) for locating *URL citations* were combined. The method applied—for the PDF version of an article from the Journal of Chemical Sciences (as shown below)—, matches (a) hyperlinks to the article if the title or URL address of the article appears in the link anchor and (b) the title or URL of the article in the text of other Web pages, even if it is not hyperlinked. We used *-site:* in order to exclude Web/URL citations from the same journal Web site. For very general short article titles we added author(s) or the journal name to our syntax to avoid retrieving unwanted results.

"Enantioselective solvent-free Robinson annulation reactions" OR

[www.ias.ac.in/chemsci/pdf-Jun2001/Pc3049.pdf](http://www.ias.ac.in/chemsci/pdf-Jun2001/Pc3049.pdf)  
site: ias.ac.in/

For articles available in HTML and PDF format, both URLs were combined in the searches through the Google OR operator. No previous study has used this kind of data collection method, and it is clearly more comprehensive than either of the two methods that it combines. Note that the method does not retrieve hyperlinks to articles where neither the URL of the page nor the title of the article is mentioned in the text of the page. For the latter, it is not possible to incorporate a link search into our existing searches (as shown above), because Google does not allow its link search command to be combined with any other search types.

Although our initial manual checking of Google searches showed that using OR operators could expand the individual search results, by retrieving Web sources such that either the title or the URL of the OA articles would appear in the link

anchor or text, inconsistencies in the functionality of the Google OR operator mean that this approach does not work perfectly (Bar-Ilan, 2005; Notes, 2003), but it works well enough for our purposes.

We excluded very long URLs (e.g., for articles stored in databases) from a few journals (11 out of 108), because we found that adding them to a query either reduced the total search results from our title searches or produced no hits. Most of these journals were from SciELO (<http://www.scielo.br>), an electronic library covering a collection of Brazilian scientific journals. We selected the option “repeat the search with the omitted results included” (if it was displayed) to retrieve the total number of results in Google. Note that all the ISI, Google Scholar, and Google Web/URL searches in this study were conducted for each discipline during the relatively short period of September-October 2005 in order to minimize the potential impact of time on increasing the number of citation counts and of variations in Google’s Web coverage.

### Google Unique and Total Web/URL Citations

We found that selecting the option “Repeat the search with the omitted results included” at the bottom of Google searches sometimes retrieved many results with similar contents from individual sites. For example, phrase searching the title of the journal article “*Micronuclei frequency in lymphocytes of individuals occupationally exposed to pesticides*” in Google retrieved 13 “unique” (one Web/URL citation per site) and 222 total results (omitted results included). Manual checking of all 222 results suggested that there were only 14 unique Web/URL citation creation motivations. In other words, the Web/URL citation motivations for the rest of the 208 results duplicated the 14 unique ones. We identified

similar patterns in a sample of 100 searches, suggesting that unique Web/citation counts are likely to be a better quantitative measure for the correlation study. In fact, Google total results often contain a separate hit for the main entry, the abstract, the PDF file, and, if available, the HTML file of each article. Although URLs of such hits are slightly different, they direct the users to the same article in the site. We believe that these results should be considered redundant. This is similar to the alternative document model concept used in link analysis (Thelwall, 2004).

In summary, for the purpose of this study, we defined Google unique Web/URL citation counts as the number of Web/URL citations, one per source site. Because Google often gives two hits per site, this number was manually calculated based upon including only one result per site. However, sometimes the same site may reappear in Google search results on different pages. The number of unique Web/URL citations was calculated for convenience by omitting the indented Google results, which reduces rather than eliminates repeated results from the same site. Because this is a relatively new issue, we decided to record both kinds of Web/URL citation counts and to investigate which one has a higher correlation with ISI citation patterns at the individual article and journal levels. The large differences between mean and median Google unique and total Web/URL citations can be seen in Table 2.

## Results

The correlation tests in Table 2 were calculated for each discipline using individual sampled articles as the unit of data collection and data analysis. This part of the study presents a broad view of disciplinary differences between

TABLE 2. Correlations between ISI, Google Scholar and Google Web/URL citation counts to OA articles and descriptive statistics for each studied discipline.

Articles sampled	Total Google Web/URL citations: Mean, median	Google unique Web/URL citations: Mean (% of ISI), median	Google Scholar citations: Mean (% of ISI), median	ISI citations: Mean, median	ISI and total Google web/URL citations	ISI and Google unique web/URL citations	ISI and Google Scholar citations	
325	41.046 22	11.492 (202%) 8	4.855 (85%) 2	5.701 3	0.259**	0.325**	<b>0.825**</b>	Biology
325	9.160 5	3.670 (155%) 3	1.027 (43%) 0	2.375 1	0.285**	0.324**	0.553**	Chemistry
325	20.604 11.5	6.901 (181%) 5	3.742 (98%) 1	3.816 1	0.415**	0.441**	0.672**	Physics
183	99.907 49	21.486 (394%) 15	10.978 (201%) 2	5.453 1	<b>0.702**</b>	<b>0.746**</b>	0.815**	Computing
185	47.940 28	19.129 (3440%) 12	2.821 (507%) 1	0.556 0	0.412**	0.438**	0.551**	Education
134	74.776 45.5	22.522 (3382%) 15	5.119 (769%) 2	0.666 0	0.467**	0.524**	0.574**	Economics
70	78.557 27.5	19.171 (1082%) 13	3.885 (219%) 1	1.771 0	0.519**	0.616**	0.766**	Sociology
103	40.135 13	14.330 (1135%) 8	2.524 (200%) 1	1.262 0	0.055	0.054	0.563**	Psychology

Note. \*\*Significant at the  $p = 0.01$  level.

counts of traditional and Web-extracted citations. Following Vaughan and Shaw (2003), Pearson correlation tests were performed if the frequency distributions were not very skewed; otherwise, Spearman correlation tests were applied. Additional results, including citations broken down by journal, are available at [http://cybermetrics.wlv.ac.uk/articledata/Kousha\\_1.doc](http://cybermetrics.wlv.ac.uk/articledata/Kousha_1.doc).

### *ISI Versus Google Scholar and Web/URL Citations*

As shown in Table 2, there is a significant correlation between the ISI and Google Scholar citation counts in all studied disciplines ( $p < 0.01$ ). Although the correlations for science disciplines are higher than for social science disciplines, especially for biology and computer science, the exceptions are sociology (high) and chemistry (low). A possible explanation for the differences in correlation coefficients is better coverage of science journals by the ISI and Google Scholar than social science journals, which will be discussed in the next section. Nevertheless, this is reasonable evidence that scholarly OA journal articles with more citations in the ISI database also have more citations reported by Google Scholar for science and the social sciences, although there may be disciplines that we have not studied for which this is not true. Moreover, because both ISI and Google Scholar citation counts measure the same aspect of formal patterns of scholarly communication (conventional citations), this is also reasonable evidence that they are essentially assessing something very similar, which might for convenience be called the intellectual impact of the work (although the exact meaning of citation counts is far more complex than this; Moed, 2005).

We can see significant, albeit weaker, correlations between ISI and Google unique citation counts as well as Google total citation counts for seven disciplines (excluding psychology) at the  $p = 0.01$  level. In both cases, there are higher correlations for computer science and sociology and lower correlations for biology and chemistry. The weaker Web citations are expected because of the more general sources of citations in Google than in Google scholar. The higher Google Scholar correlations for science disciplines are not reflected in the Web/URL citation correlations; three social science disciplines (sociology, economics, and education) have higher Web correlations (Google unique and total Web/URL citations) than three of the hard science disciplines (biology, chemistry, and physics).

Finally, there are higher correlations between ISI citations and Google unique Web/URL citations than Google total Web/URL citations in both science and social science disciplines (excluding psychology), supporting our above-mentioned argument that unique (one per site) Web/URL citation counts in Google search results are a better scholarly measure than total Web/URL citations.

### *Disciplinary Differences*

Table 2 shows that the mean number of ISI citations to sampled articles in three science disciplines (biology, chemistry, and physics) is higher than Google Scholar citations.

Moreover, the median of ISI citations to sampled articles in biology and chemistry is higher than Google Scholar citations and physics has an equal median for both kinds of citation counts. In contrast, in four social science disciplines (sociology, economics, psychology, and education), the mean and median number of ISI citations is much lower than Google Scholar citations, suggesting that Google Scholar may have particularly good coverage of sources for citations in the social sciences but slightly weak for the sciences. In computer science, the mean and median number of Google Scholar citations is much higher than ISI citations (about double). One explanation is that in computer science there is an existing established Web citation database (CiteSeer/ResearchIndex); in computer science, conferences are very important and their proceedings are frequently made available online, where Google Scholar may find them. In physics, conferences are not the major dissemination mechanism but open access publishing of preprints is common, via ArXiv.org. Table 2 also shows that the mean and median number of Google Web/URL citations (unique and total) is much higher than both ISI citations. This corroborates the work of Vaughan and Shaw (2005).

### **Further Analysis**

In this section we use additional data to investigate further the main research questions.

### *Overlaps Between ISI and Google Scholar Citations*

We analyzed the overlaps between ISI citations and Google Scholar citations in the four scientific disciplines in order to find out whether they were using the same data or whether their coverage was significantly different. We did not include social science disciplines because there were too few ISI-indexed journals at the time of this study. We sampled 880 research articles from 39 Open Access ISI-indexed biology, physics, chemistry, and computing journals in 2001. We manually checked 4,184 ISI citations against 5,589 Google Scholar citations in order to identify common citations. We calculated the percentage of relative overlap (RO) for both the ISI and Google Scholar in each discipline (Table 3). The overall relative overlap percentage for the ISI was 57% (2,387). In other words, out of 4,184 ISI citations, 2,387 records were duplicated in Google Scholar. This overlap percentage was relatively higher in biology (66%), physics (62%), and computing (57%) and considerably lower in chemistry (33%). There were only small differences between the mean and median ISI and Google Scholar citations in biology and physics. However, in the case of chemistry, mean and median ISI citations were considerably higher than in Google Scholar. One possible explanation for the low number of Google Scholar citations in chemistry is missing data from individual large publishers. For instance, is the considerable difference in chemistry related to those *Elsevier* publications that are not covered by Google Scholar (Jacso, 2004)? In fact, we found that none of the ISI citations



TABLE 3. Overlaps between ISI and Google Scholar citations to open access ISI-indexed journals (2001) in four science disciplines.

Discipline	Journals	Articles sampled	ISI citations: Mean, median	Google Scholar citations: Mean, median	ISI overlapping with Google Scholar citations: Mean, median	ISI relative overlap (overlapping ISI citations divided by total ISI citations)	Google Scholar relative overlap (overlapping ISI citations divided by total Google Scholar citations)	Google unique citations (not overlapping with the ISI citations)	Unique ISI citations (not overlapping with Google Scholar citations)
Biology	14	262	4.9 3	4.2 2	3.2 2	66%	76%	266	441
Chemistry	10	276	2.4 1	1 0	0.78 0	33%	78%	61	450
Physics	11	262	4.2 2	5.0 1.5	2.6 1	62%	52%	623	421
Computing	4	82	13.6 5	35.2 13	7.7 3	57%	22%	2252	485
Total	39	882	4184	5589	2387	57%	43%	3202	1797

TABLE 4. Increases in ISI and Google Scholar citations to OA journal articles within four months in four science disciplines.

Discipline	ISI citations (Oct. 2005)	ISI citations (Jan. 2006)	ISI citation increase	Google Scholar citations (Oct. 2005)	Google Scholar citations (Jan. 2006)	Google Scholar citation increase
Biology	1208	1288	6.2%	923	1113	17.1%
Chemistry	456	668	31.7%	228	279	18.3%
Physics	1061	1111	4.5%	1097	1313	16.5%
Computing	962	1117	13.9%	2125	2884	26.3%
Total	3687	4184	11.9%	4373	5589	21.7%

from journals published by the American Chemical Society were in Google Scholar, providing an alternative simple explanation.

#### *Changes in ISI and Google Scholar Citation Counts*

We examined the increase in both ISI and Google Scholar citations to the same sampled articles within a four-month period from October 2005 to January 2006 (Table 4). Our purposes were to validate the Google Scholar citation database (because it is relatively new) as well as to compare the percentage increase in both citation indexes during the same time period. We found that in the four science subject areas the percentage increase for ISI citations was about 12%; however, for Google Scholar this increase was about 22%. Although the percentage increase for Google Scholar citations was considerably higher than for the ISI counterpart in biology, physics, and computing, it was significantly lower for chemistry. A possible explanation is that the Google Scholar has increased its coverage in three of the subjects, but not chemistry, perhaps because it was unable to negotiate new access to chemistry publishers' records. This difference suggests that the absolute number of citations would be more volatile in Google Scholar, perhaps because it can add entire digital libraries at once, whereas the ISI's quality control procedures naturally make for more gradual changes to its databases.

#### *Application: Journal-Level Citation Counting*

If Google Scholar or the Web are to be used for impact factor calculations, then it is important to compare citation scores for each journal separately. A strong association would support the use of Google Scholar or the Web as an alternative source of journal citation impact. Correlations tests were thus performed using individual journals for each of the 108 journals (rather than articles as the unit of data analysis in each selected discipline): between the average number of ISI citations and the average number of Google Scholar citations and the average number of Web/URL citations (Table 5). For each variable, we calculated the total number of citations (ISI, Google Scholar, and Web/URL citations) a journal received divided by the number of articles in the sample set. As shown in Table 5, there is a highly significant correlation between the average number of ISI citations and the average number of Google Scholar citations in all the disciplines at the  $p = 0.01$  level. Thus, it seems that OA journals having higher average ISI citations also have higher average Google Scholar citations. Hence, Google Scholar is a promising tool for measuring the intellectual impact of OA journals as an alternative to conventional citation indexes in a wide variety of disciplines.

We also found significant correlations between the average number of ISI citations and the average number of Google unique Web/URL citations (as we previously defined) at the journal level for each discipline (Table 5), but



TABLE 5. Correlations between average ISI and average Google Scholar and Google Web citation counts to OA journals.

Journals	Average ISI vs. average Google total web/URL citations	Average ISI vs. average Google unique web/URL citations	Average ISI vs. average Google Scholar citations	
21	0.248	0.622**	0.938**	Biology
15	0.443	0.721**	0.744**	Chemistry
16	0.503*	0.547*	0.926**	Physics
12	0.782**	0.831**	0.880**	Computer
17	0.597*	0.611**	0.782**	Education
11	0.689*	0.734*	0.655**	Economics
16	0.470	0.540*	0.789**	Sociology/Psychology

Note. \*Significant at the  $p < 0.05$  level.

\*\*Significant at the  $p < 0.01$  level.

lower than the correlations between average ISI and Google Scholar citations (Table 2). We found significant correlations between average ISI and average Google total Web/URL citations at the journal level for four disciplines. The results provide additional evidence that there is a stronger relationship between average ISI citations and average unique Google Web/URL citations to OA journals than Google total Web/URL citations.

## Conclusions

### *Google Scholar*

In answer to the first research question, we found a significant correlation between Google Scholar citations and ISI citations in all disciplines. This is strong evidence that Google Scholar has a widely applicable value in citation counting. Nevertheless, see the limitations below.

### *Web/URL Citation*

In answer to the second research question, we found a significant correlation between Google Scholar citations and ISI citations in all disciplines except psychology. This is evidence that Web/URL citations are more problematic than Google Scholar citations and should be used cautiously. Psychology must be further investigated to assess whether it is a pathological case or an example of a common exception.

### *Conventional Versus Web Citations*

We found a relatively stronger relationship between the average ISI and Google Scholar citations than Google Web/URL citations in nearly all cases at the article and journal level. A likely explanation is that both ISI and Google Scholar are measuring formal scholarly patterns equivalent to formal citations and that Web/URL citations include types of informal citation in addition to the formal ones. It is not clear, however, which is the better type of measure for research impact. For example, if many Web/URL citations represented genuine uses of research (e.g., in education or industry) then this could be seen as desirable, whereas if most Web/URL citations were in replicated library lists, then this could be problematic. In addition, there are disciplinary

differences in research that lead to varied emphasis on things like electronic publication, books, journals, and conferences (Kling & McKim, 1999; Whitley, 2000) and variations in usage patterns for similar electronic resources, including e-journals (e.g., Fry & Talja, 2004). It may be that the Web contains objects of value in the social sciences, such as course reading lists, that would not be used or valued in the sciences where research is often less directly tied to teaching. It would be interesting to compare peer judgments to assess whether articles that were relatively highly cited online were overrated or genuinely useful.

### *Unique Web/URL Citation Versus Total Web/URL Citation*

In terms of the Web/URL citation method, separating out the unique and total Web/URL citation counts in our Google search results produced relatively higher correlations between the average ISI and Google unique Web/URL citations, suggesting that counting a maximum of one match per site produces improved results. The fact that Google total results retrieves many multiple links that direct the users to the same place in a site can be considered as the main reason for the relatively stronger relationship between ISI and Google unique Web/URL citation counts in all our correlation tests at the journal and article level.

Our new Web/URL citation method is, in theory, more comprehensive than either Web citation or URL citation. Nevertheless, it has some practical drawbacks. In our version, it requires some manual labor to ensure that only one citation per Web site is allowed, although this could be automated through the Google API (Mayr & Tosques, 2005). Moreover, the method required modification for some journals because the URLs were too long, which creates a potential unfairness. Finally, it is not yet clear that the most comprehensive solution is the best, especially if many of the additional citations may come from undesired sources.

### *Disciplinary Differences Between Web and Conventional Citation Patterns*

In answer to the third research question, there are clear disciplinary differences between conventional and Web-based citation patterns. Why is there a relatively stronger correlation between ISI and Google Scholar citation counts

in science disciplines than social science at the journal and article level? It may be relevant that about 77% (49 of 64) of selected journals in science disciplines and only 13% (6 of 44) of social science journals were indexed by the ISI at the time of this study. The descriptive statistics for ISI and Google Scholar citation counts (Table 2) show that in three pure science disciplines (biology, chemistry, and physics, but excluding computer science) the distribution of citations is relatively less skewed than in social science. Consequently, it may be that higher coverage of citation information in both ISI and Google Scholar is the important factor for higher commonality between citation patterns in science disciplines. This speculation is supported by the fact that, in contrast to the three pure science disciplines, in the four social science subject areas the mean and median Google Scholar citation counts are remarkably higher than ISI citations (Table 2). This suggests that Google Scholar is a more comprehensive tool for citation tracking for social science. However, the quality of sources of citations (citing documents) retrieved by Google Scholar is an important factor to take into account (as for Web/URL citations), and future research must address this complex issue.

### Limitations

The methods here, and particularly Web/URL citation counting give an advantage to Web-published sources of citation, and so it would be unfair to use them to compare non-OA journals—even though (non-OA) Web publishing seems to be standard now for the major academic publishers, and some believe that OA adoption will be widespread soon (Shadbolt, Brody, Carr, & Harnad, 2006). The disciplinary differences in the extent of overlap between ISI and Google Scholar citations, and particularly the apparent differing rates of increase of citations, should also provide caution for those seeking to use Google Scholar citations as evidence of scientific impact. Note also that our disciplines were effectively a convenience sample, self-selected by volume of OA journal use and that there were technical problems with some of the queries. Nevertheless, it is reasonable to use statistics derived from Google Scholar or Web/URL citations (cautiously) for impact calculations when ISI data is not available or the results are intended to be indicative rather than evaluative. An overall problem that time may solve, however, is that the relatively low number of OA journals in 2001 meant that many disciplines not studied in this article did not have enough OA journals to analyze. Finally, it would be interesting to see a similar study to this one applied to non-English language journals, perhaps to help assess the extent of any language bias in the ISI databases.

### References

About Google Scholar. (2005). Retrieved December 12, 2005, from <http://scholar.google.com/scholar/about.html>

Almind, T.C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to "Webometrics." *Journal of Documentation*, 53(4), 404–426.

Antelman, K. (2004). Do Open-Access articles have a greater research impact? *College & Research Libraries*, 65(5), 372–382.

Bar-Ilan, J. (1999). Search engine results over time—A case study on search engine stability. *Cybermetrics* 2/3. Retrieved January 26, 2006, from <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>

Bar-Ilan, J. (2004). The use of Web search engines in information science research. *Annual Review of Information Science and Technology*, 38, 231–288.

Bar-Ilan, J. (2005). Expectations versus reality—Search engine features needed for Web research. *Cybermetrics*, 9(1). Retrieved July 21, 2006, from <http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html>

Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11(9). Retrieved December 23, 2005, from <http://www.dlib.org/dlib/september05/bauer/09bauer.html>

Belew, R.K. (2005). Scientific impact quantity and quality: Analysis of two sources of bibliographic data. Retrieved April 28, 2006 from <http://arxiv.org/abs/cs.LR/0504036>

Björneborn, L., & Ingwersen, P. (2001). Perspectives of Webometrics. *Scientometrics*, 50(1), 65–82.

Borgman, C., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, pp. 3–72. Medford, NJ: Information Today Inc.

Brin, S., & Page, L. (1998). The anatomy of a large scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117.

Brody, T., Stamerjohanns, H., Vallières, F., Harnad, S., Gingras, Y., & Oppenheim, C. (2004). The effect of open access on citation impact. Retrieved November 13, 2001, from <http://www.ecs.soton.ac.uk/~harnad/Temp/OA-TAadvantage.pdf>

Egghe, L. (2000). New informetric aspects of the Internet: Some reflections—many problems. *Journal of Information Science*, 26(5), 329–335.

Friend, F. (2006). Google Scholar: Potentially good for users of academic information. *The Journal of Electronic Publishing*. Retrieved April 28, 2006, from <http://hdl.handle.net/2027/spo.3336451.0009.105>

Fry, J. (2004). The cultural shaping of ICTs within academic fields: Corpus-based linguistics as a case study. *Literary and Linguistic Computing*, 19(3), 303–319.

Fry, J., & Talja, S. (2004). The cultural shaping of scholarly communication: Explaining e-journal use within and across academic fields. In *Proceedings of the 67th ASIST Annual Meeting 2004* (Vol. 41, pp. 20–30). Medford, NJ: Information Today.

Glänzel, W. (2003, February). On some on some principle differences between citations and citation links. A methodological and mathematical approach. NERDI lecture delivered at NIWI, KNAW, Amsterdam. (Updated version of a paper presented at the 6th Nordic Workshop on Bibliometrics, Stockholm, October 4–5, 2001.)

Goodrum, A.A., McCain, K.W., Lawrence, S., & Giles, C.L. (2001). Scholarly publishing in the Internet age: A citation analysis of computer science literature. *Information Processing & Management*, 37(5), 661–676.

Harnad, S. (1990). Scholarly skywriting and the prepublication continuum of scientific inquiry. *Psychological Science* 1, 342–343. Retrieved November 12, 2004, from <http://www.cogsci.soton.ac.uk/~harnad/Articles/Harnad/harnad90.skywriting.html>

Harnad, S. (1991). Post-Gutenberg galaxy: The fourth revolution in the means of production of knowledge. *Public-Access Computer Systems Review*, 2(1), 39–53. Retrieved November 12, 2004, from <http://www.cogsci.soton.ac.uk/~harnad/Articles/Harnad/harnad91.postgutenberg.html>

Harnad, S., (1999). The future of scholarly skywriting, in the sky: Visions of the information future. Retrieved November 12, 2004, from <http://cogprints.org/1698/00/harnad99.aslib.html>

Harnad, S., & Brody, T. (2004). Comparing the impact of open access (oa) vs. non-oa articles in the same journals. *D-Lib Magazine*, 10(6). Retrieved May 2, 2006, from <http://www.dlib.org/dlib/june04/harnad/06harnad.html>

Harnad, S., & Carr, L. (2000). Integrating, navigating, and analysing open eprint archives through open citation linking (the OpCit project). *Current Science*, 79(5), 629–638.

Harter, S.P. (1996). The impact of electronic journals on scholarly communication: A citation analysis. *The Public-Access Computer Systems*

- Review, 7. Retrieved November 13, 2001, from <http://info.lib.uh.edu/pr/v7/n5/hart7n5.html>
- Harter, S., & Ford, C. (2000). Web-based analysis of E-journal impact: Approaches, problems, and issues. *Journal of the American Society for Information Science*, 51(13), 1159–76.
- Herring, S.D. (2002). Use of electronic resources in scholarly electronic journals: A citation analysis. *College and Research Libraries*, 63(4), 334–340.
- Ingwersen, P. (1998). The calculation of Web impact factors. *Journal of Documentation*, 54(2), 236–243. ISI press release essay on the impact of open access journals: A citation study from Thomson Scientific. Retrieved November 13, 2004, from <http://www.ISInet.com/oaj>
- Jacso, P. (2004). Google Scholar Beta. Péter's Digital Reference Shelf. Retrieved Jan 10, 2006, from <http://snipurl.com/dwco>
- Jacso, P. (2005a). Google Scholar: The pros and the cons. *Online Information Review*, 29(2), 208–214.
- Jacso, P. (2005b). As we may search: Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537–1547. Retrieved April 28, 2006, from <http://www.ias.ac.in/currsci/nov102005/1537.pdf>
- Kim, H.J. (2000). Motivations for hyperlinking in scholarly electronic articles: A qualitative study. *Journal of the American Society for Information Science*, 51(10), 887–899.
- Kling, R., & Callahan, E. (2003). Electronic journals, the internet, and scholarly publishing. *Annual Review of Information Science and Technology*, 37, 127–177.
- Kling, R., & McKim, G. (1999). Scholarly communication and the continuum of electronic publishing. *Journal of American Society for Information Science*, 50(10), 890–906.
- Kousha, K., & Thelwall, M. (2006). Motivations for URL citations to open access LIS library and information science articles: Exploring characteristics of sources of Web citation. *Scientometrics*, 68(3), 501–517.
- Kurtz, M.J. (2004). Restrictive access policies cut readership of electronic research journal articles by a factor of two. Harvard-Smithsonian Centre for Astrophysics, Cambridge, MA. Retrieved November 13, 2001, from <http://opcit.eprints.org/feb190a/kurtz.pdf>
- Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., & Murray, S.S. (2005). Worldwide use and impact of the NASA Astrophysics Data System digital library. *Journal of the American Society for Information Science & Technology*, 56(1), 36–45.
- Lawrence, S. (2001). Free online availability substantially increases a article's impact. *Nature*, 411, 521. Retrieved November 13, 2001, from <http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>
- Lawrence, S., & Giles, C.L. (1999). Accessibility of information on the Web. *Nature*, 400, 107–109.
- Mayr, P., & Tosques, F. (2005). Google Web APIs: An instrument for Webometric analyses? Retrieved January 20, 2006, from [http://www.ib.hu-berlin.de/%7Emayr/arbeiten/ISSI2005\\_Mayr\\_Toques.pdf](http://www.ib.hu-berlin.de/%7Emayr/arbeiten/ISSI2005_Mayr_Toques.pdf)
- Metrop, W., & Nieuwenhuysen, P. (2001). Internet search engines-fluctuations in document accessibility. *Journal of Documentation*, 57(5), 623–651.
- Moed, H.F. (2005). *Citation analysis in research evaluation*. New York: Springer.
- Notess, G. (2003). Google inconsistencies. Search Engine Showdown. Retrieved July 21, 2006, from <http://www.searchengineshowdown.com/features/google/inconsistent.shtml>
- Notess, G.R. (2005). Scholarly Web searching: Google Scholar and Scirus. *Online*, 29(4).
- Oppenheim, C. (2000). Do patent citations count? In B. Cronin & H.B. Atkins (Eds.), *The Web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 405–432). Metford, NJ. Information Today Inc ASS Monograph Series.
- Pauly, D., & Stergiou, K. (2005). Equivalence of results from two citation Thomson ISI's Citation Index and Google's Scholar service. *Ethics in Science and Environmental Policies*. December 22, 33–35.
- Rousseau, R. (1997). Situations: An exploratory study. *Cybermetrics*, 1(1). Retrieved November 14, 2001, from <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>
- Shadbolt, N., Brody, T., Carr, L., & Harnad, S. (2006). The open research Web: A preview of the optimal and the inevitable. In N. Jacobs (Ed.), *Open access: Key strategic, technical and economic aspects*. Chandos. Retrieved May 3, 2006, from <http://cogprints.org/4841/02/shad-bch.pdf>
- Shin, E.-J. (2003). Do Impact Factors change with a change of medium? A comparison of impact factors when publication is by article and through parallel publishing. *Journal of Information Science*, 29(6), 527–533.
- Smith, A.G. (1999). A tale of two Web spaces: Comparing sites using Web impact factors. *Journal of Documentation*, 55(5), 577–592.
- Smith, A.G. (2004). Web links as analogues of citations. *Information Research*, 9(4). Retrieved March 20, 2005, from <http://informationr.net/ir/9-4/article188.html>
- Swan, A., & Brown, S. (2004). Report of the JISC/OSI open access journal authors survey, 1–76. Retrieved April 20, 2006, from [http://www.jisc.ac.uk/uploaded\\_documents/JISCOreport1.pdf](http://www.jisc.ac.uk/uploaded_documents/JISCOreport1.pdf)
- Swan, A., & Brown, S. (2005). Open access self-archiving: An author study, 1–97. Retrieved April 20, 2006, from <http://eprints.ecs.soton.ac.uk/10999/01/jisc2.pdf>
- Thelwall, M. (2004). *Link analysis: An information science approach*. San Diego: Academic Press.
- Thelwall, M. (2006). Interpreting social science link analysis research: A theoretical framework. *Journal of the American Society for Information Science and Technology*, 57(1), 60–68.
- Thelwall, M., Vaughan, L., & Björneborn, L. (2005). *Webometrics. Annual Review of Information Science and Technology*, 39, 81–135. Medford, NJ: Information Today Inc.
- Van Impe, S., & Rousseau, R. (2006). Web-to-print citations and the humanities. *Information—Wissenschaft und Praxis*, 57(8), 422–426.
- Vaughan, L. (2004). New measurements for search engine evaluation proposed and tested. *Information Processing & Management*, 40(4), 677–691.
- Vaughan, L., & Hysen, K. (2002). Relationship between links to journal Web sites and Impact Factors. *Aslib Proceedings, New Information Perspectives*, 54(6), 356–361.
- Vaughan, L., & Shaw, D. (2003). Bibliographic and Web citations: What is the difference? *Journal of the American Society for Information Science and Technology*, 54(14), 1313–1324.
- Vaughan, L., & Shaw, D. (2005). Web citation data for impact assessment: A comparison of four science disciplines. *Journal of the American Society for Information Science and Technology*, 56(10), 1075–1087.
- Vaughan, L., & Thelwall, M. (2003). Scholarly use of the Web: What are the key inducers of links to journal Web sites? *Journal of the American Society for Information Science and Technology*, 54(1), 29–38.
- Whitley, R. (2000). *The intellectual and social organization of the sciences* (2nd ed.). Oxford: Oxford University Press.
- Zhao, D. (2005). Challenges of scholarly publications on the Web to the evaluation of science—A comparison of author visibility on the Web and in print journals. *Information Processing and Management*, 41(6), 1403–1418.
- Zhao, D., & Logan, E. (2002). Citation analysis using scientific publications on the Web as data source: A case study in the XML research area. *Scientometrics*, 54(3), 449–472.