



## Using the Web for research evaluation: The Integrated Online Impact indicator

Kayvan Kousha<sup>a,\*</sup>, Mike Thelwall<sup>b</sup>, Somayeh Rezaie<sup>c</sup>

<sup>a</sup> Department of Library and Information Science, University of Tehran, Jalle-Ahmad Highway, P.O. Box 14155-6456, Tehran, Iran

<sup>b</sup> School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1ST, UK

<sup>c</sup> Department of Library and Information Science, Shahid Beheshti University, Evin Ave, P.O. Box 1983963113, Tehran, Iran

### ARTICLE INFO

#### Article history:

Received 4 August 2009

Received in revised form 6 October 2009

Accepted 12 October 2009

#### Keywords:

Web citation

Online impact

Research evaluation

Webometrics

### ABSTRACT

Previous research has shown that citation data from different types of Web sources can potentially be used for research evaluation. Here we introduce a new combined Integrated Online Impact (IOI) indicator. For a case study, we selected research articles published in the *Journal of the American Society for Information Science & Technology (JASIST)* and *Scientometrics* in 2003. We compared the citation counts from Web of Science (WoS) and Scopus with five online sources of citation data including Google Scholar, Google Books, Google Blogs, PowerPoint presentations and course reading lists. The mean and median IOI was nearly twice as high as both WoS and Scopus, confirming that online citations are sufficiently numerous to be useful for the impact assessment of research. We also found significant correlations between conventional and online impact indicators, confirming that both assess something similar in scholarly communication. Further analysis showed that the overall percentage for unique Google Scholar citations outside the WoS were 73% and 60% for the articles published in JASIST and Scientometrics, respectively. An important conclusion is that in subject areas where wider types of intellectual impact indicators outside the WoS and Scopus databases are needed for research evaluation, IOI can be used to help monitor research performance.

© 2009 Elsevier Ltd. All rights reserved.

### 1. Introduction

Journal-based citation databases have long been used for the impact assessment of research. However, in some disciplines, especially social sciences and the humanities, journal citations can sometimes be insufficient and broader types of impact indicators (e.g., from books, monographs and conference papers) may also be helpful for research assessment (see Moed, 2005; Nederhof, 2006). Moreover, there are informal scholarly activities that may also influence scholarly work (Becher & Trowler, 2001; Crane, 1972), but cannot be measured using traditional citation analysis tools and techniques.

Previous studies have indicated that the Web contains relevant information for research evaluation, especially in the social sciences (Kousha & Thelwall, 2007b). Several studies have assessed the value of different Web sources for impact assessment through *Google* (e.g., Kousha & Thelwall, 2007a; Vaughan & Shaw, 2003, 2005), *Google Scholar* (e.g., Harzing & van der Wal, 2009; Jacsó, 2005a, 2005b; Kousha & Thelwall, 2008b; Meho & Yang, 2007) and more recently *Google Books* (Kousha & Thelwall, 2009). Moreover, some investigations have examined more specialized informal Web sources which

\* Corresponding author. Tel.: +98 21 22255694; fax: +98 21 22255694.

E-mail addresses: [kkoosha@ut.ac.ir](mailto:kkoosha@ut.ac.ir) (K. Kousha), [m.thelwall@wlv.ac.uk](mailto:m.thelwall@wlv.ac.uk) (M. Thelwall), [s.aries80@yahoo.com](mailto:s.aries80@yahoo.com) (S. Rezaie).

might be helpful for intellectual impact assessment such as *presentations* (Thelwall & Kousha, 2008) and *course reading lists* (Kousha & Thelwall, 2008a).

The above studies have used online impact indicators independently, but no investigation has used a *range of specific types* of Web sources for intellectual impact assessment. In this study we introduce the IOI indicator to extract citation data from broader types of research-relevant Web sources. The underlying goal is to develop alternative methods for scholars or research institutions, authors, journal editors, and academic publishers to use Web sources for additional citations to their work. Moreover, scientometricians may also use IOI when conducting research evaluation studies in subject areas in which WoS or Scopus citations are neither comprehensive nor sufficient for impact assessment (e.g., the arts and humanities).

A case study was conducted on research articles published in 2003 in two library and information science journals: *JASIST* and *Scientometrics*. We used WoS, Scopus, Google Scholar and Google Books to extract formal citations to the selected journal articles. We also applied different techniques to extract informal citations located in online syllabuses, PowerPoint presentations and blogs.

## 2. Related studies

There is now a considerable body of literature concerning online impact indicators for research evaluation (see below). The primary aim of the most related studies is to examine whether the Web can be used for intellectual impact indicators as an additional source for conventional citation analysis (e.g., WoS/Scopus citations).

### 2.1. Formal evidence of online impact

Online formal impact is defined in this study as that which is derived from citations found within the reference sections of online documents, either from full text documents or cross-reference and Web-based citation indexes. Two major databases, Google Scholar and Google Books, are amongst those that can be used to locate such citations to journal articles.

#### 2.1.1. Google Scholar

Google Scholar has become a powerful online citation analysis tool. Google Scholar claims to include some documents that would not be indexed by WoS and Scopus, such as preprints, conference papers, and articles from academic publishers, professional societies, universities and other scholarly organizations (About Google Scholar, 2009). Bauer and Bakkalbasi (2005) compared citations from WoS, Scopus and Google Scholar targeting articles from *JASIST* published in 1985 and 2000. They found that for articles published in 2000, Google Scholar provided significantly higher citation counts than either WoS or Scopus. However, there were no significant differences between WoS and Scopus for the studied years. In contrast, Belew (2005) selected six academics at random and compared citations to publications by these authors indexed by WoS with those reported by Google Scholar, finding a small overlap between the two databases.

Jacsó (2005a) conducted several early test searches of WoS, Scopus and Google Scholar and compared their citing sources. He found that WoS and Scopus “have almost identical citedness scores” for the highly cited papers of *Current Science*. However, the coverage of Google Scholar was “abysmal”. In another study, Jacsó (2005b) assessed the citedness scores in WoS and Google Scholar for breadth of source coverage and the ability of the software to identify the cited documents correctly. For papers published in 22 volumes of the *Asian Pacific Journal of Allergy and Immunology* he found that the aggregate citedness score was 1355 for the 675 papers retrieved by WoS, and 595 for 680 papers found in Google Scholar. He concluded that “the poor capabilities of GS to consolidate the matching records inflates both the number of hits and the citedness score”.

Although earlier studies have shown the limited coverage of Google Scholar vs. WoS and Scopus (see above) as well as “inflated citation counts” in Google Scholar, partly due to “the inclusion of non-scholarly sources” (Jacsó, 2006), more recent studies have given more promising results. Walters (2006), for example, compared the contents of Google Scholar and seven other databases (Academic Search Elite, AgeLine, ArticleFirst, GEOBASE, POPLINE, Social Sciences Abstracts, and Social Sciences Citation Index) within the multidisciplinary subject area of later-life migration. Each database was evaluated using a set of 155 core articles selected in advance—the most important studies of later-life migration published from 1990 to 2000. Among the eight databases, Google Scholar provided the most comprehensive citation coverage of core articles about later-life migration. It covered 27% more core articles than SSCI and at least twice as many as each of the three disciplinary databases (AgeLine, POPLINE, and GEOBASE). The results suggested that Google Scholar indexes not just a large number of documents, but a large number of high-quality research papers, at least in some fields.

Meho and Yang (2007) also compared citations from WoS with Scopus and Google Scholar to examine the impact of adding Scopus and Google Scholar citation counts to the ranking of LIS faculty (1996–2005). Using citations to the work of 25 LIS faculty members as a case study, they found that Google Scholar provides greater coverage of conference proceedings as well as non-English language journals. More importantly, they revealed that the overlap between WoS and Scopus was 58.2%. However, the overlap between Google Scholar and the union of WoS and Scopus was low (30.8%), indicating that there is “high uniqueness between the three tools”. Kousha and Thelwall (2007a) found significant correlations between WoS and Google Scholar citations for a sample of 1650 articles from 108 Open Access journals published in 2001 in four science and four social science disciplines, indicating that both measure similar formal scholarly patterns to formal citations. However, there were large disciplinary differences, suggesting that Google Scholar gives more comprehensive results in the social

sciences than in science (excluding computing). They also analysed the overlaps between WoS and Google Scholar citations for four science disciplines in order to find out whether they were using the same or significantly different data. This overlap percentage was higher in biology (66%), physics (62%) and computing (57%), and lower in chemistry (33%), indicating that there are clear disciplinary differences. The percentage increase within a 4-month period was relatively large: about 12% for WoS citations and 22% for Google Scholar. Norris and Oppenheim (2007) compared holdings and citation records of WoS, Scopus, CSA Illumina and Google Scholar against selected sets of articles in social sciences. They found that whilst Google Scholar has the highest citation counts, “when these are examined individually, it is clear that the results have not, at the very least, been de-duplicated”, validating Jacsó’s (2005b) claims.

Other studies using Google Scholar citations have been motivated by the h-index. Bar-Ilan (2008) compared the h-indexes of a list of 40 highly cited Israeli researchers based on citation counts retrieved from the WoS, Scopus and Google Scholar, finding the number of Google Scholar citations to be considerably higher than the WoS and Scopus for mathematicians and computer scientists, but lower for high-energy physicists. Similarly, Harzing and van der Wal (2008) compared Google Scholar-based metrics with the Journal Impact Factors to assess journal impact (h-index, g-index and the number of citations per paper). They found strong correlations between Google Scholar metrics and the traditional Journal Impact Factors. They concluded that the use of Google Scholar generally results in “a more comprehensive and possibly more accurate measure of true impact” in the studied subject areas. In another study, Harzing and van der Wal (2009) compared the Google Scholar h-index and the Journal Impact Factor for a sample of 838 journals in economics and business, finding that the Google Scholar provides a more accurate and comprehensive measure of journal impact. They concluded that the Google Scholar h-index should be considered as a supplement to traditional citation-based impact analyses.

In contrast to the above studies which report higher numbers of Google Scholar citations than WoS citations in several social sciences and humanities disciplines, other studies have shown that in chemistry (Bornmann, Marx, Schier, Rahm, Thor, & Daniel, 2009; Kousha & Thelwall, 2008b) and high-energy physics (Bar-Ilan, 2008) lower citation counts are returned by Google Scholar compared to the fee-based databases such as WoS and Scopus. As mentioned above, in chemistry one explanation for such missing citations in Google Scholar is that it could not directly access the major journal publishers such as the American Chemical Society (ACS) in order to build citing and cited associations (see Bornmann et al., 2009; Kousha & Thelwall, 2008b).

### 2.1.2. Google Books

Although books are a key scholarly platform in many social science and humanities disciplines, traditional bibliometric methods and tools have failed to include citations from books and monographs for impact assessment. For the first time, Kousha and Thelwall (2009) used Google Books and compared citations from books to journal articles in a total of ten science, social science and humanities disciplines. Book citations reached 31–212% of WoS citations and hence were numerous enough to supplement WoS citations in the social sciences and humanities covered, but not in the sciences (3–5%) except for computing (46%). These results suggest that Google Books is a valuable new source of citation data for the social sciences and humanities.

## 2.2. Informal evidence of online impact

We conceive online informal impact as that which is derived from any online sources that are a by-product of any type of scholarly use of papers, indicating that the research has been found useful. In contrast to formal impact, in which citations are explicitly mentioned in reference sections of academic documents, informal impact is associated with wider scholarly communication activities (e.g., discussion, correspondence and teaching). For instance, citations in academic course reading lists or syllabuses suggests that the cited works were useful enough to be mentioned for teaching purposes and hence could be evidence of educational impact.

### 2.2.1. Educational impact

Kousha and Thelwall (2008a) used a new method to assess the extent to which citations from online syllabuses could be a source of evidence about the educational utility of research. An analysis of online syllabus citations to 70,700 articles published in 2003 in the journals of 12 subjects indicated that online syllabus citations were numerous enough to be useful in some social sciences, but not in science. These results suggest that online syllabus citations provide a valuable additional source for impact assessment of research, especially in the social sciences and humanities.

### 2.2.2. Online presentations

Many conference or seminar presentations are available online (e.g., as PowerPoint presentations) and they can also be a promising source for impact assessment. Thelwall and Kousha (2008) used Google to assess how WoS journals in ten science and ten social science disciplines were cited in online PowerPoint presentations. Although few journals were cited frequently enough in online PowerPoint files to make impact assessment worthwhile, the method can still be used as additional evidence for the impact of academic research in the context of a combined indicator or multiple indicators.

### 3. Research questions

The main objective of this paper is to assess the online impact of research articles published by *JASIST* and *Scientometrics* in 2003 as a case study in the use of the IOI. The project uses different Web sources and methods for potential inclusion in impact assessment which might be useful for scholars, authors or journals. Hence, the method applied here can indicate how published academic research which cannot be traced through WoS and Scopus can be used to assess scholarly output. The following specific questions drive the research.

1. How are research articles published in both *JASIST* and *Scientometrics* in 2003 cited online by Web sources that can be used for intellectual impact assessment?
2. Comparing both WoS and Scopus citations, is the IOI method sufficient and useful for research impact assessment?
3. Does IOI correlate with their conventional counterparts (WoS/Scopus)?

### 4. Methods and procedures

In the current study different methods were used to capture and analyse intellectual impact indicators. We selected *JASIST* and *Scientometrics* as a case study because they are among the most influential journals in the information science and had similar citation impact as reported in 2003 Journal Citation Reports (1.473 and 1.251, respectively). Another important reason for selecting the above journals was to investigate how journals' specialization in a very specific subject area may affect the extent of informal impact evidence that can be found online (such as teaching, popularization or scholarly correspondence or discussions). *Scientometrics* mainly covers studies of science from a quantitative perspective whereas *JASIST* has much broader coverage of information science and technology. We chose the year 2003 to give sufficient time (about 6 years) for articles to attract citations by different Web sources such as non-WoS journal articles, conference papers, books, dissertations and even by informal sources of scholarly impact such as course reading lists, presentation files and blogs. All the searches in this study were conducted during the relatively short period of April–May 2009 in order to reduce the impact of time on increasing the number of citations.

Note that of 95 *JASIST* papers indexed in the WoS 10 were not found in Scopus at the time of this study. Hence, we applied an alternative method to record the number of citations to them. We searched the exact article titles in the main Scopus search interface and selected the "References" field. This method retrieved citations to the missing *JASIST* papers in the references of other documents indexed by Scopus. We manually checked the cited references to remove false matches. We used a similar approach for the single *Scientometrics* article absent from Scopus.

Although previous studies have reported that many WoS-indexed articles and citing sources in the field of chemistry were not in Google Scholar (see Bornmann et al., 2009; Kousha & Thelwall, 2008b), we located all the selected articles in from both *JASIST* and *Scientometrics*. This extensive overlap between WoS and Google Scholar coverage may be the result of agreements between Google Scholar and individual publishers. Hence, we think that Wiley and Springer, which are the publishers for *JASIST* and *Scientometrics* respectively, have opened up their academic journals to Google Scholar indexing. In contrast, in chemistry Google Scholar has been unable to directly access American Chemical Society publications, including several high impact chemical science titles (see Meho & Yang, 2007, p. 2109; Kousha & Thelwall, 2008b, p. 288). Moreover, a study of Walters (2006) on Google Scholar coverage of the multidisciplinary area of later-life migration indicated that Google Scholar "covers 27% more core articles than SSCI" (Walters, 2006, p. 1125). This is discussed again in the conclusions.

#### 4.1. WoS and Scopus citations

We used WoS advanced search option to record the number of citations from more than 10,000 of the highest impact scientific research journals. In order to mine cited references to research articles published by *JASIST* and *Scientometrics*, we searched the above journal titles in the "publication name" field separately and limited the results to 2003 and articles (omitting reports, editorials, and book reviews). The results were exported to a spreadsheet for further analysis. Note that we also used the Scopus database which has broader coverage (about 17,000 journal titles) than WoS. We used a similar approach to the above for Scopus citations. Ultimately, we selected 95 and 72 research articles published by *JASIST* and *Scientometrics* respectively which were indexed in both WoS and Scopus.

#### 4.2. Formal online impact

For the purpose of this study, 'formal online impact' covers citations to journal articles in the reference sections of online scholarly documents (e.g., dissertations, conference papers, books, research reports, preprints and post-prints). This scope is similar to formal scholarly communication (Borgman & Furner, 2002), research-oriented (Bar-Ilan, 2004, 2005) and research impact (Vaughan & Shaw, 2005). We used two main online sources to capture Web citations: Google Scholar and Google Books.

#### 4.2.1. Google Scholar citations

Google Scholar (<http://scholar.google.com>) claims to include peer-reviewed articles, theses, books, abstracts and articles from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations. For Google Scholar citation counts, we searched the titles of all research articles published in *JASIST* and *Scientometrics* in 2003 in the main Google Scholar search page as phrase search and recorded the number of Google Scholar citations by clicking the “cited by” option below each retrieved record. Because Google Scholar sometimes retrieves duplicate records for the same journal articles we manually checked the search results against the original citation information to avoid duplicates.

#### 4.2.2. Google Book citations

Google Books (<http://books.google.com>) provides full text searching that can be used to locate citations in books. For Google Book citations, we searched the titles of selected articles as phrase searches in the advanced Google Book search option and selected “all books” to retrieve citations to articles from books (usually with an ISBN) (for the method see Kousha & Thelwall, 2009)

### 4.3. Informal online impact

In contrast to formal impact, which is commonly based upon journal citation counting, informal impact comprises a wider range of academic sources, such as course reading lists, conference presentations and scholarly correspondence or discussions. Moreover, the value of these types of indirect intellectual impact indicators may vary between disciplines. We used the following three informal online sources.

#### 4.3.1. Educational impact

Educational impact assessment suggests how much an academic work has influenced teaching. In particular, online syllabuses can be used to identify research that lecturers recommend for their students (e.g., recommended, additional, required, suggested readings). We used the previously applied method to assess when articles were mentioned in online course reading lists. We manually searched the exact article titles in the Google main interface and combined them with (syllabus OR “reading list”). We manually checked the retrieved results to remove possible false matches (for methods see Kousha & Thelwall, 2008a).

#### 4.3.2. Scholarly presentation impact

Every year, thousands of presentations are given at conferences, workshops, seminars and other scholarly events worldwide. Presentations play an important communication role, often reporting research for the first time to other scholars in the same subject area. As for educational impact assessment (see above), it is possible to assess how research papers are used for scholarly presentations. For this purpose, we again used Google and searched for online PowerPoint files mentioning any articles selected for this study. We again conducted manual data checking to remove false results (see Thelwall & Kousha, 2008).

#### 4.3.3. Blog impact

There is an increasing number of blogs in different scientific disciplines (e.g., chemistry, education, physics, library and information science). In some cases, messages posted to blogs may convey a degree of intellectual impact. For instance, scholars may post a message which mentions an article to back up a scholarly discussion, to give background information or as a recommendation to other people. We think that these can also be considered as informal evidence of research impact, although the recognition of this type of intellectual impact may be a very subjective and complicated issue. We used Google Blogs (<http://blogsearch.google.com>) as a practical tool to identify messages posted to blogs and indexed, mentioning any articles selected for scholarly related reasons. We again conducted manual checking to remove false results.

## 5. Findings

### 5.1. WoS/Scopus vs. IOI citations

Table 1 reports citation counts of research articles published in *JASIST* and *Scientometrics* in 2003 based upon the WoS and Scopus databases. It shows that the mean and median number of WoS citations to *JASIST* articles are higher than to *Scientometrics*. However, the mean and the median of WoS citations to research articles published in both journals are lower than Scopus citations, confirming that Scopus has larger coverage of journal articles.

Table 2 reports citations to research articles published in *JASIST* and *Scientometrics* based on different online impact sources including Google Scholar, Google Books, Google Blogs, course reading lists and presentations. As defined above, the total number of the above online citations is the IOI indicator. Table 2 indicates that the IOI mean and median for *JASIST* articles are 25 and 16 whereas the IOI mean and median for *Scientometrics* articles are 16 and 11 respectively. The results suggest that *JASIST* papers tended to attract more online intellectual impact than *Scientometrics*. As shown in Table 2, *JASIST* attracted more citations from course reading lists (38). However we couldn't find any citations to *Scientometrics* articles

**Table 1**

Citation counts for research articles published in JASIST and Scientometrics in 2003 based on WoS and Scopus searches.

Sources of citation	JASIST			Scientometrics		
	No. of papers	Citation count	Mean Median	No. of papers	Citation count	Mean Median
Web of Science	95	1051	11.06 6	72	577	8.01 5
Scopus	95	1237	13.02 7	72	599	8.31 5

**Table 2**

Online indicators applied for impact assessment of research articles published in JASIST and Scientometrics.

Online sources of citation	JASIST		Scientometrics	
	Citation count	Mean Median	Citation count	Mean Median
Google Scholar	2109	22.2 15	1030	14.30 9.5
Google Books	217	2.28 1	78	1.08 1
Course Reading Lists	38	0.4 0	0	0 0
Google Blogs	33	0.34 0	32	0.44 0
PowerPoint Presentation	22	0.23 0	9	0.12 0
Total	2419	25.45 16	1149	15.94 10.5

from online syllabuses, suggesting a better overall educational utility for JASIST papers. The possible explanations for this surprising result will be discussed again in the conclusions.

Google Books, Google Blogs and PowerPoint files for both journals are compared in Table 2. Most importantly, Table 2 shows that the mean and median IOI are twice as high as both WoS and Scopus citations (see Table 1). However, the quality of online impact is unknown and it would be unfair to directly compare WoS and Scopus citations to online impact indicators. Hence, our results about online impact calculations are intended to be indicative rather than evaluative and should be used cautiously for research evaluation.

Table 3 shows the proportion of WoS and Scopus citations per IOI citation and vice versa for both JASIST and Scientometrics. For instance, the first column of Table 3 reports that the ratio of WoS citations to IOI is 0.43. In other words, IOI citations are 2.3 times more numerous than WoS citations.

## 5.2. Article-level analysis

As shown in Tables 4 and 5, there are significant correlations at the article level between the conventional (WoS and Scopus) and online impact indicators, suggesting that they measure similar or related aspects of scholarly communication. For instance, the second column of Table 4 reports that there are strong correlations (0.876\*\* for JASIST and 0.842\*\* for Scientometrics) between WoS citation counts and the IOI Indicator for research articles published in 2003 in both JASIST and Scientometrics. The third column also reports correlation between WoS citation counts and online impact indicators excluding Google Scholar citations. We excluded Google Scholar to remove possible overlapping citations from journal

**Table 3**

The proportion of WoS and Scopus citations per IOI citation.

JASIST		Scientometrics	
WoS/IOI	Scopus/IOI	WoS/IOI	Scopus/IOI
0.43	0.51	0.50	0.52
JASIST		Scientometrics	
IOI/WoS	IOI/Scopus	IOI/WoS	IOI/Scopus
2.30	1.96	1.99	1.92

**Table 4**

Correlations between WoS citations and online impact indicators.

	WoS and IOI (GS, GB, PPT, edu, blog)	WoS and IOI (excluding GS)	WoS and informal (excluding GS and GB)
JASIST ( <i>n</i> = 95)	0.876**	0.724**	0.464**
Scientometrics ( <i>n</i> = 72)	0.842**	0.558**	0.391**

\*\*Significant at the *p* = 0.01 level.**Table 5**

Correlations between Scopus citations and online impact indicators.

	Scopus and IOI (GS, GB, PPT, edu, blog)	Scopus and IOI (excluding GS)	Scopus and informal (excluding GS and GB)
JASIST	0.915**	0.730**	0.460**
Scientometrics	0.629**	0.879**	0.233**

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

articles in WoS and Google Scholar and to examine how this might influence the correlation pattern. As shown in the third column of **Table 4**, there are also significant correlations between WoS citations and unique online citation impact (excluding Google Scholar), suggesting that even when we excluded online citations mainly from journal and conference papers, there is still a significant relationship between WoS and online citations appearing in books, PowerPoint presentations, online course reading lists and blogs. The last column of **Table 4** reports relatively low but significant correlations between WoS citation counts and online “informal” impact measures from PowerPoint presentations, online course reading lists and blogs (excluding citations in journals and books). Hence, the result suggests that formal and informal impact indicators used in this study assess something similar.

Note that the Spearman correlation test was performed instead of Pearson because all the citation frequency distributions were skewed.

**Table 5** also reports correlations between the Scopus citations and different online impact indicators, indicating that there are strong relationships between variables. Most notably, **Table 5** indicates that there are higher correlations between Scopus citations and online impact indicators than WoS citations excluding informal online impact indicators, suggesting that Scopus has a wider coverage of different type of citing sources.

**Appendices A and B** list the top 20 articles in *JASIST* and *Scientometrics* based on the IOI. It shows that the most cited articles based on IOI also attracted high WoS and Scopus citations. The last column of **Appendix A** reports the ratio of IOI citations to WoS, indicating that the proportion of IOI citations to WoS citations ranges from 1.2 to 6.7. As shown in **Appendix A**, in several cases papers had a moderate impact according to the WoS and Scopus citations, but a very high impact according to IOI. For instance, Kling’s article “*A bit more to it: Scholarly communication forums as socio-technical interaction networks*” (fourth row in **Appendix A**) received 19 WoS citations, but at least 90 citations from different sources of online impact including 69 citations from Google Scholar, 17 from books indexed by Google Books, 2 in course reading lists and 2 citations in PowerPoint presentations. In other words, the IOI impact of the above article seems 4.7 times higher than WoS citations would suggest. **Appendix B** shows that the proportions of IOI citations to WoS citations for *Scientometrics* articles are considerably lower than to *JASIST* papers, ranging from 1.2 to 3.3, suggesting that they tended to attract less interest from the online sources used this study.

### 5.3. Overlapping and unique WoS and GS citations

We analysed the common citations between WoS and Google Scholar for the sample of 30 *JASIST* papers and 20 *Scientometrics* articles in order to estimate the percentage of relative overlap (RO) as well as unique citations for both WoS and Google Scholar (**Table 6**). We manually checked 262 WoS citations against 811 Google Scholar citations to 30 sampled *JASIST* papers, finding 221 common citations. Consequently, the overall relative overlap percentage for WoS was about 84% and for Google Scholar was about 27%. The last column of **Table 6** reports that the relative unique citations percentage for WoS is about 16% and for Google Scholar is about 73%, indicating that there are many unique online citations to *JASIST* papers which were not indexed by WoS and so they would be “invisible” in any impact assessment studies using WoS. As shown in **Table 6**, similar citation patterns can be observed for both the relative overlap and unique citations for the sampled *Scientometrics* articles. An important corollary is that in subject areas where wider types of intellectual impact indicators are need for research evaluation, Google Scholar can be used as additional source of citation data.

### 5.4. Different IOI weights

The IOI calculation is the simple sum of the separately calculated citations but there is no reason why the IOI should not be a weighted indicator. In other words the choice of *a*, *b*, *c*, *d* and *e* in the following is 1 in this article but in theory any other values could be used.

$$\text{IOI} = a \times \text{Scholar} + b \times \text{PowerPoint} + c \times \text{Blogs} + d \times \text{Syllabus} + e \times \text{Books}$$

**Table 6**  
The relative overlap and unique citations for both WoS and Google Scholar.

Journals	Articles sampled	Citation count		Overlapping citations	Relative overlap		Unique citations	
		WoS	GS		WoS	GS	WoS	GS
JASIST	30	262	811	221	84.35% (221)	27.25% (221)	15.65% (41)	72.75% (590)
Scientometrics	20	178	337	135	75.84% (135)	40.05% (135)	24.16% (43)	59.95% (202)



Without a “gold standard” for the value of each paper or another means of deciding the relative importance of the publication types it is not possible to determine the optimal weights. Nevertheless, it intuitively seems that Google Scholar and Book citations should be at least more valuable than Blog citations.

One way to seek an optimal set of weights is to use WoS citations as a gold standard and select weights that make the IOI results closest to WoS citations using linear regression. This approach was attempted but gave insignificant coefficients  $b$ – $e$ . In other words, PowerPoint, Blogs, Syllabus, and Books do not provide significantly better predictor of WoS citations than Scholar citations alone. This is despite the fact that all of the five types of IOI citations above correlate (Spearman) significantly with WoS citations for both JASIST and Scientometrics, with the exception of Syllabus and PowerPoint citations for Scientometrics. The regression approach may fail because WoS citations miss types of impact that the others measure and hence it is not a good gold standard. This suggests that other means should be used to judge the relative value of the five sources, such as an expert panel.

## 6. Limitations

Although citation impact assessment based upon online syllabuses or reading lists could be a valuable source of evidence about the educational utility of research, the results of Google syllabus searches in this study do not include all online syllabuses or teaching materials deposited online by the instructors or academic staff. For instance, there is restricted access to the content of many online academic syllabuses and course reading lists (e.g., in password protected databases) perhaps for copyright reasons. Hence, our syllabus searches only include the portion of the teaching materials that has been crawled by Google.

Another issue is that we used simple sum of citation counts for different types of online source. However, the weights for online intellectual impact indicators in this study could easily be different. For instance, online citations in academic course readings lists may reflect a higher degree of potential intellectual impact than Web citations created for scholarly publicity and discussion. As for traditional citation analysis in which new measures are proposed for citation impact (e.g., [Ajiferuke & Wolfram, 2009](#); [Hirsch, 2005](#)), further investigation is needed to examine the way online citations may be counted and weighted.

In this study the IOI results were based on manual checking to guarantee that online citations were created for scholarly reasons. Therefore, the IOI citation extraction method here cannot be used for automatic impact assessment of research. Moreover, we only selected two journals in the library and information science discipline; hence it is not known whether disciplinary differences would influence the extent of online impact indicators. Our online impact assessment included citations in the selected databases; hence less is known about the potential value of other sources of online impact (i.e., Web CVs, scientific databases) in research evaluation.

## 7. Conclusions

We found that IOI citations are numerous enough to make a difference in research intellectual impact assessment and could be used as additional source for monitoring research performance. Although WoS citations are widely used for impact assessment, the IOI method applied in this study is a novel practical approach to extract citation data from wider online sources. Hence, the IOI method can reveal how journal articles receive citations from a range of online scholarly Web documents such as conference papers, dissertations, e-prints, research reports and others which were previously impossible to trace through conventional serial-based citation databases. Moreover, we identified new sources of informal intellectual impact (discussion messages and blogs) that have not been used in previous Scientometrics studies, but which may reflect informal scholarly uses of research and may be useful in some social sciences, art and humanities disciplines in which scholarly discussions are valued.

The mean and median IOI were twice as high as both WoS and Scopus citations, but the quality of online impact is unknown and it would be unfair to directly compare WoS and Scopus citations to online impact indicators. Consequently, online impact calculations should be used cautiously for research evaluation and they should not be used as a replacement for the conventional impact indicators. However, in some social sciences, art and humanities disciplines which may depend mainly on information published in books, monographs, conference papers, and technical reports, IOI may be particularly valuable for ranking universities or scholars.

There were 38 citations to JASIST articles from online syllabuses and reading lists, but no citations to Scientometrics articles from the same sources. This suggests that articles in an individual very specific field of science may have little direct teaching impact or have an impact that is not much wider than the field itself. In support of this, a recent study of inter-journal citations found Scientometrics to receive an unusually high proportion of citations from articles published in the same journal ([Schneider, 2009](#)). However, the subject analysis of the 38 citations to JASIST papers showed that 9 out of 38 citations (24%) were from papers about scientometrics or Webometrics which are relevant to Scientometrics. Information storage and retrieval (47%) and information use and seeking behaviour (16%) were other topics of the JASIST papers which attracted citations from the online reading lists. The result supports the previous finding that “the articles most cited by Web syllabuses or reading lists tend to be reasonably highly WoS-cited, but that the converse is not true: some highly WoS-cited articles appear never to be mentioned in Web syllabuses or course reading lists” ([Kousha & Thelwall, 2008a, p.](#)

2065). Another reason for finding no syllabus citation to Scientometrics papers might that there are relatively few academic courses about Scientometrics, Informetrics and Webometrics worldwide (however see the limitation above), or that the journal Scientometrics is less often available to students, as a more specialist journal that less universities subscribe to.

## Acknowledgments

The authors would like to thank anonymous reviewers for their valuable comments on earlier drafts of this paper.

## Appendix A. Top 20 cited JASIST articles (2003) based on IOI.

First author	Article	WoS	Scopus	GS	GB	Reading	Blog	ppt	IOI	IOI/WoS
Ahlgren, P	Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient	68	65	106	4	0	0	0	110	1.62
Borlund, P	The concept of relevance in IR	56	63	91	10	7	0	0	108	1.93
<b>Kling, R</b>	<b>A bit more to it: Scholarly communication forums as socio-technical interaction networks</b>	<b>19</b>	<b>31</b>	<b>69</b>	<b>17</b>	<b>2</b>	<b>0</b>	<b>2</b>	<b>90</b>	<b>4.74</b>
White, HD	Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists	49	45	80	6	2	1	0	89	1.82
Vaughan, L.	Scholarly use of the Web: What are the key inducers of links to journal Web sites?	53	60	87	2	0	0	0	89	1.68
Hoad, TC	Methods for identifying versioned and plagiarized documents	20	36	80	3	0	0	2	85	4.25
Song, D	Towards context sensitive information inference	13	22	78	6	0	0	0	84	6.46
Thelwall, M	The connection between the research of a university and counts of links to its Web pages: An investigation based upon a classification of the relationships of pages to the research of the host university	27	30	61	5	1	0	1	68	2.52
Vaughan, L	Bibliographic and Web citations: What is the difference?	38	48	63	3	0	0	0	66	1.74
Meho, LI	Modeling the information-seeking behaviour of social scientists: Ellis's study revisited	23	29	44	7	3	2	0	56	2.43
Wang, PL	Mining longitudinal Web queries: Trends and patterns	34	44	47	6	1	0	0	54	1.59
Hara, N	An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration	24	30	45	4	2	0	0	51	2.13
Morris, SA	Time line visualization of research fronts	23	23	49	0	0	1	1	51	2.22
Heinz, S	Efficient single-pass index construction for text databases	7	12	38	6	1	1	1	47	6.71
White, HD	Author cocitation analysis and Pearson's r	37	34	44	2	0	0	0	46	1.24
Morillo, F	Interdisciplinarity in science: A tentative typology of disciplines and research areas	25	24	39	5	0	1	1	46	1.84
Thelwall, M	Graph structure in three national academic Webs: Power laws with anomalies	17	22	39	4	0	1	1	45	2.65
Chen, HC	HelpfulMed: Intelligent searching for medical information over the Internet	21	20	36	7	0	0	0	43	2.05
Small, H	Paradigms, citations, and maps of science: A personal history	19	23	35	5	2	0	0	42	2.21
Choi, Y	Searching for images: The analysis of users' queries for image retrieval in American history	15	22	32	4	4	1	1	42	2.80
Mothe, J	DocCube: Multi-dimensional visualisation and exploration of large document sets	11	12	36	4	0	0	0	40	3.64

Table is ranked based on IOI citations in the tenth column. GS=Google Scholar citations; GB=Google Books citations; Reading=citations from course reading lists; Blog=Google Blogs citations; ppt=citations from Google PowerPoint presentations; IOI=Integrated Online Impact indicators.

**Appendix B. Top 20 cited Scientometrics articles (2003) based on IOI.**

First author	Article	WoS	Scop	GS	GB	Read	Blog	ppt	IOI	IOI/WoS
Aksnes, DW	A macro study of self-citation	40	39	78	3	0	0	0	81	2.03
Hullmann, A	Publications and patents in nanotechnology—an overview of previous studies and the state of the art	31	35	53	5	0	0	1	59	1.90
Glanzel, W	A new classification scheme of science fields and subfields designed for scientometric evaluation purposes	37	32	50	1	0	2	0	53	1.43
Thelwall, M	Linguistic patterns of academic Web use in Western Europe	31	35	46	3	0	2	1	52	1.68
Meyer, M	Towards hybrid Triple Helix indicators: A study of university-related patents and a survey of academic inventors	23	18	42	3	0	1	0	46	2.00
Leydesdorff, L	The mutual information of university-industry-government relations: An indicator of the Triple Helix dynamics	18	18	39	3	0	3	0	45	2.50
Heimeriks, G	Mapping communication and collaboration in heterogeneous research networks	31	30	41	2	0	0	0	43	1.39
Ranga, LM	Entrepreneurial universities and the dynamics of academic knowledge production: a case study of basic vs. applied research in Belgium	11	11	34	3	0	0	0	37	3.36
Zitt, M	Correcting glasses help fair comparisons in international science landscape: Country indicators as a function of ISI database delineation	18	16	31	4	0	0	0	35	1.94
Glanzel, W	Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon	19	16	29	2	0	2	0	33	1.74
Stegmann, J	Hypothesis generation guided by co-word clustering	13	13	20	4	0	1	1	26	2.00
Hood, WW	Informetric studies using databases: opportunities and challenges	20	18	21	2	0	2	0	25	1.25
Glanzel, W	Patents cited in the scientific literature: an exploratory study of 'reverse' citation relations	11	12	22	3	0	0	0	25	2.27
Bhattacharya, s	Characterizing intellectual spaces between science and technology	16	14	19	2	0	2	0	23	1.44
Kyvik, S	Changing trends in publishing behaviour among university faculty, 1980–2000	12	15	21	2	0	0	0	23	1.92
Van Looy, B	Do science-technology interactions pay off when developing technology? An exploratory investigation of 10 science-intensive technology domains	13	13	20	0	0	1	0	21	1.62
Braun, T	A quantitative view on the coming of age of interdisciplinarity in the sciences 1980–1999	8	7	18	2	0	1	0	21	2.63
Inonu, E	The influence of cultural factors on scientific production	11	10	20	0	0	0	0	20	1.82
Ugolini, D	The visibility of Italian journals	11	13	15	0	0	1	0	16	1.45
Danell, R	Regional R&D activities and interactions in the Swedish Triple Helix	8	11	14	1	0	1	0	16	2.00

Table is ranked based on IOI citations in the tenth column.

**References**

- About Google Scholar. (2009). Retrieved November 7, 2009, from <http://scholar.google.com/scholar/about.html>.
- Ajiferuke, I., & Wolfram, D. (2009). Citer analysis as a measure of research impact: Library and Information Science as a Case Study. In B. Larsen, & J. Leta (Eds.), *Proceedings of the 12th International Conference of the International Society for Scientometrics and Informetrics* (pp. 798–808).
- Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., & Daniel, H. (2009). Convergent validity of bibliometric Google Scholar data in the field of chemistry—citation counts for papers that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using Google Scholar, Science Citation Index Scopus, and Chemical Abstracts. *Journal of Informetrics*, 3(1), 27–35.
- Bar-Ilan, J. (2004). A microscopic link analysis of universities within a country—The case of Israel. *Scientometrics*, 59(3), 391–403.
- Bar-Ilan, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257–271.
- Bar-Ilan, J. (2005). What do we know about links and linking? A framework for studying links in academic environments. *Information Processing & Management*, 41(4), 973–986.

- Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11(9). Retrieved December 1, 2008, from <http://www.dlib.org/dlib/september05/bauer/09bauer.html>.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories* (2nd ed.). Milton Keynes, UK: Open University Press.
- Belew, R. (2005). Scientific impact quantity and quality: Analysis of two sources of bibliographic data. Retrieved May 21, 2006, from <http://arxiv.org/abs/cs.IR/0504036>.
- Borgman, C., & Furner, J. (2002). Scholarly communication and bibliometrics. In B. Cronin (Ed.), *Annual Review of Information Science and Technology* 36 (pp. 3–72). Medford, NJ: Information Today Inc.
- Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago: University of Chicago Press.
- Harzing, A., & van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, 8(1), 61–73. Retrieved September 27, 2009 from <http://www.int-res.com/articles/esep2008/8/e008p061.pdf>.
- Harzing, A., & van der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology*, 60(1), 41–46.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 16569–16572.
- Jacsó, P. (2005a). 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 September 27, 2009, from <http://www.ias.ac.in/currsci/nov102005/1537.pdf>.
- Jacsó, P. (2005b). Comparison and analysis of the citedness scores in Web of Science and Google Scholar. *Lecture Notes in Computer Science*, 3815, 360–369. Retrieved September 27, 2009, from <http://www.jacso.info/PDFs/jacso-comparison-analysis-of-citedness.pdf>.
- Jacsó, P. (2006). Deflated, inflated and phantom citation counts. *Online Information Review*, 30(3), 297–309. Retrieved September 20, 2009, from <http://www.jacso.info/PDFs/jacso-deflated-inflated.pdf>.
- Kousha, K., & Thelwall, M. (2007a). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055–1065.
- Kousha, K., & Thelwall, M. (2007b). The Web impact of open access social science research. *Library and Information Science Research*, 29(4), 495–507.
- Kousha, K., & Thelwall, M. (2008a). Assessing the impact of research on teaching: An automatic analysis of online syllabuses in science and social sciences. *Journal of the American Society of Information Science and Technology*, 59(13), 2060–2069.
- Kousha, K., & Thelwall, M. (2008b). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics*, 74(2), 273–294.
- Kousha, K., & Thelwall, M. (2009). Google Book search: Citation analysis for social science and the humanities. *Journal of the American Society of Information Science and Technology*, 60(8), 1537–1549.
- Meho, L., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. New York: Springer.
- Nederhof, A. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81–100.
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics*, 1(2), 161–169.
- Schneider, J. W. (2009). Mapping of cross-reference activity between journals by use of multidimensional unfolding: Implications for mapping studies. In B. Larsen, & J. Leta (Eds.), *Proceedings of ISSI 2009* (pp. 443–454). Rio, Brazil: BIREME/PAHO/WHO and Federal University of Rio de Janeiro.
- Thelwall, M., & Kousha, K. (2008). Online presentations as a source of scientific impact? An analysis of PowerPoint files citing academic journals. *Journal of the American Society of Information Science and Technology*, 59(5), 805–815.
- 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.
- Walters, W. H. (2006). Google Scholar coverage of a multidisciplinary field. *Information Processing & Management*, 43(4), 1121–1132.