MEASURING THE IMPACT OF KNOWLEDGE
A Comparison of Web of Science and Google Scholar

John Mingers and Lea Lipitakis
Kent Business School, University of kent, Canterbury CT7 2PE, U.K.

Keywords: Citations, Impact metrics, Research benchmarks.

Abstract: Assessing the quality of the knowledge produced by management academics is increasing being metricated. Moreover, emphasis is being placed on the impact of the research rather than simply where it is published. The main metric for impact is the number of citations a paper receives. Traditionally this data has come from the ISI Web of Science but research has shown that this has poor coverage in the social sciences. A newer and different source for citations is Google Scholar. In this paper we compare the two on a dataset of over 1200 publications from a UK Business School. The results show that Web of Science is indeed poor in the area of management and that Google Scholar, whilst somewhat unreliable, has a much better coverage.

1 INTRODUCTION

Assessing researchers' productivity and the impact of knowledge generated is increasingly being metricated and the number of citations is one of the main measures that is used. This occurs at an individual level in promotion and hiring decisions and increasingly at an institutional level in evaluating whole departments and universities. In the UK, the Research Excellence Framework (REF) intends to use citation analysis along with peer review in future decisions about the allocation of research funding. There are many complex issues involved in using metrics for this purpose and the Higher Education Funding Council for England (HEFCE) has commissioned several reports and is currently undertaking a pilot exercise.

One of the major problems, especially in the social sciences, is the source of the citations. The primary database has conventionally been Thompson's ISI Web of Science (WoS) which records all citations from papers in about 8,700 journals. Whilst this coverage is reasonable in many of the sciences it is acknowledged to be limited in social science, partly because many journals are not included and partly because much research is published in books and conferences which are not covered at all. In recent years alternatives have been developed that work in a similar manner, e.g., Scopus, but one of the main rivals is Google Scholar (GS). This works in a different fashion by searching the internet and other digital repositories to find citations in a wide range of sources.

Several studies have compared the two sources in general (Jacso, 2005), and in particular disciplines (Bakkalbasi, Bauer, Glover, & Wang, 2006; Bar-Ilan, 2008; Meho & Yang, 2007), while HEFCE's commissioned reports have concentrated mainly on the sciences because of the known problems in the social science. Their pilot exercise, for example, includes almost no social science subjects (HEFCE, 2008b). No one that we are aware of has looked specifically at the management literature. So, the purpose of this paper is to investigate the extent to which WoS and GS do in fact record research outputs and citations in business and management, and to discover whether there are any particular patterns in their coverage or lack of it. To do this we have taken all the publications of academics at a UK Business School over the period 2001-2007, together with a selection from earlier years, and processed them through WoS and GS. Our results are reported after a review of the relevant literature and a description of our methodology. The School is representative in that it covers all the main business disciplines and is in the top 30 in the UK.

2 WEB OF SCIENCE AND GOOGLE SCHOLAR

The Web of Science. This covers over 8,700 primarily

Mingers J. and Lipitakis L. (2009).
MEASURING THE IMPACT OF KNOWLEDGE - A Comparison of Web of Science and Google Scholar.
In *Proceedings of the International Conference on Knowledge Management and Information Sharing*, pages 112-116
DOI: 10.5220/0002278401120116

English-language journals out of approximately 22,500 listed in Ulrich's Periodicals Directory. It does not include reports, books or conference proceedings (although proceedings are just beginning to be incorporated in 2008). WoS records every paper published in these journals together with their citations and then allows access in a variety of ways including citation reports on journals and individual authors.

In recent years a range of alternative databases have emerged, some discipline specific such as the ACM Digital Library and some generic such as Elsevier's Scopus. These are of three types: those that involve searching the full text of the document for citations where the text may be contained in the database (e.g., Emerald full text or Scirus) or may be home pages and repositories on the web (e.g., Google Scholar); those that allow the user to search the cited reference field of the document (e.g., EBSCO products); and finally those like WoS that are primarily designed for capturing citations (e.g., Scopus). Several studies have been carried out comparing these different sources often in different disciplines and Meho and Yang (2007) provide a good overview.

In this study we limit ourselves to comparing WoS with GS specifically in the discipline of Business and Management. The two databases have very different modes of operation. WoS has a clearly specified list of journals and records all the citations from those journals. Its coverage is generally considered to be good in many of the natural sciences but poor in the social sciences and humanities (HEFCE, 2008a; Mahdi, D'Este, & Neely, 2008; Moed & Visser, 2008). It has tools that help with the unique identification of authors – one of the major problems in collecting accurate citations. In contrast, GS has a scope and reliability that is in general unknown (Harzing & van der Wal, 2008; Jacso, 2008). It searches web pages and also has access to the websites of certain publishers but the exact details remain secret. The results generally have a wide coverage but can include many works that are not specifically research oriented, e.g., teaching notes, discussions and reports. It is relatively difficult to pin down a specific author, especially if they have a common name, and often the bibliographic details of the citing sources are wrong or incomplete hence getting accurate results is extremely time consuming.

Meho and Yang (2007), in their study of a Department of Library and Information Science, found that 42% of GS citations came from journals, 34% from conference papers, 10% from dissertations and theses and 14% from other sources. They found 2023 citations to their source documents (including only journal items and conference papers from 1996-2005) in WoS, 2301 in Scopus and 4181 in GS. Combining WoS and Scopus produced 2733 unique citations while including those from GS pushed the total up to 5285. Thus, WoS produced only 48% of the citations in GS, and only 38% of the citations generated by a combination of all three. Walters (2007) studied 155 core articles in the area of later-life migration across a range of citation databases. GS had the greatest coverage (93%) and WoS next best with 73%. Whilst this study did not look at citations, it did examine the range of sources used by GS in terms of publishers (sometimes a source of criticism (Tenopir, 2005)) and found no undue bias.

The Centre for Science and Technology Studies at Leiden University (CSTS) has presented several commissioned reports. In 2008 they analysed the submissions to the 2001 Research Assessment Exercise (RAE) (Moed, Visser, & Buter, 2008), looking in the main at the science subjects. They did however do some analysis across all units of assessment. Table 1 shows the coverage of outputs in WoS. We can see that economics has the best coverage with 68% of its total outputs in WoS rising to 78% of the journal papers. However, management generally has only 38% covered and accounting and finance a mere 22%. The latter result is because a significant number of high quality accounting and finance journals are not included in WoS.

Evidence Ltd (Evidence Ltd, 2004) conducted research for ESRC producing a bibliometric profile for selected disciplines including business and management, accounting and economics. The main results are also shown in Table 1. It is worrying that the two results are not particularly close. This no doubt reflects in part the difficulties of unambiguously identifying individual papers in these databases, and differing practices over what to do with ambiguous references, but it is noticeable that there is not even agreement on the total number of submitted outputs to the RAE.

The research also looked in detail at the number of cites per paper (cpp) for those papers that could be found in WoS but only for the departments graded as 4, 5 or 5* (the highest grades). The number of citations is obviously time dependent so these figures will be an average across the period of the RAE, i.e, papers published in 1995 would have five years of citations, those published in 2000 only one year. Thus economics averages 8 cites per paper but accounting and finance only 4. This is clearly related to the coverage of journals – areas with a higher coverage show greater numbers of citations.

Table 1: CSTS analysis of WoS coverage of RAE2001 outputs. Evidence Ltd figures are in brackets.

	Submitted outputs	% of outputs that are journal papers	% of outputs that are in WoS	% journal papers that are in WoS	Mean cites per paper (4-5* departments)
Economics	2,879 (3255)	86.2% (76%)	67.5% (47%)	78.3% (62%)	(8.0)
Business & Management	9,746 (9942)	81.8% (80%)	37.9% (31%)	46.3% (38%)	(6.3)
Library & Information Management	1,259	59.0%	31.7%	53.7%	
Accounting and Finance	779 (811)	85.2% (82%)	21.7% (17%)	25.5% (20%)	(3.9)

Table 2: GS and WoS citations by publication type.

Publication Type	Num.	n%	No of Pubs. in GS	% GS	No of Cites found in GS	No of Pubs. in WoS	% WoS	No of Cites found in WoS	GS Cites Per Paper	WoS Cites Per Paper
Books	19	1.57	11	57.89	405				36.82	
Book Section	109	8.99	64	58.72	479				7.48	
Conference Papers	330	27.23	154	46.67	399				2.59	
Conference Proceedings	29	2.39	14	48.28	58				4.14	
Edited Books	12	0.99	8	66.67	313				39.13	
Journal Articles	593	48.93	548	92.41	5608	292	49.24	1519	10.23	5.20
Reports	115	9.49	63	54.78	319				5.06	
Unpublished Work	4	0.33	3	75.00	18				6.00	
Web Pages	1	0.08	1	100.00	1				1.00	
TOTAL	1212	100.0	866	71.45	7600	292			8.78	

Citation rates normalised to the rates for the disciplinary field were also calculated (the "Leiden methodology (van Raan, 2003)). In this approach, results above 1.0 show that the publications are generating more citations than the average for the field. The figures for business and management were 1.47 (for 4-graded departments), 1.90 (5-graded) and 2.27 (5*-graded) showing both high impact and that the impact increases with the RAE grade. The equivalent figures for accounting are: 0.28, 0.82 and 1.07 showing that it is not simply the lack of WoS journals – accounting departments, especially at the lower end, gain relatively very few citations.

3 STUDY RESULTS

The data consisted of over 1200 research outputs produced by staff at Kent Business School from 2001 to 2007 (which is the RAE period) including some from earlier years. Each publication was individually looked up in GS and WoS (where it was a journal paper). This is a very time-consuming exercise, especially for GS, since the quality of the data is poor – there are often multiple entries for a single item because the forms of reference are inconsistent or inaccurate (Jacso, 2008).

Table 2 shows the main results. We have included all publication types even though many would not be submitted to a REF. We can see that the majority of the outputs are journal papers (50%) with the next category being conference papers (12%). Looking first at the GS coverage, we found 71.5% of all the publications including 92% of the journal papers – a very significant proportion. Surprisingly perhaps, given the high presence of publishers' websites, only 58% of books and 67% of edited books were found. Other areas of low coverage

Table 3: Citations by field or subject area.

	Papers	GS citations	WoS citations	GS cpp	WoS cpp	WoS/GS %
Agriculture, environment, natural resources	110	784	326	7.1	3.0	42%
Engineering	15	104	65	6.9	4.3	62%
Economics	61	601	194	9.9	3.2	32%
Operational research and management science	95	1508	729	15.7	7.7	49%
Applied mathematics and statistics	37	357	169	9.7	4.6	47%
Management, tourism, public sector, industrial relations	146	2287	917	15.7	6.3	40%
Social science	37	295	90	8.00	2.4	30%
Information systems and computer science	41	1332	205	32.5	5.0	15%
Business	35	159	17	4.5	0.5	11%

were conferences and reports. In contrast, WoS would only cover journal papers and only found 49% of those in the sample. This figure is similar to, although slightly higher than, those found for the RAE generally in Table 1. On some occasions the journal was apparently on the WoS list but the actual paper did not appear. This was generally found to be because the journal was not part of WoS at the time that the paper was published, sometimes because there was a gap in the journal history.

Moving to citations, GS found a considerable number for all publication types. The mean cpp were highest for books (36.8) and edited books (39.1) with the figure for journal papers being 10.2. WoS found 1,519 citations for the 292 papers it included giving a cpp of 5.2. Again, this was quite similar to the RAE result of 6.3. These citations represented only about 27% of the citations that GS found for the same database of papers. This is significantly lower than the 48% figure that Meho and Yang found.

We also looked to see if these proportions had changed over time but in both cases there were year-to-year variations but no apparent trend. It could be argued that if the purpose of using these measures is to compare departments or research centres then it doesn't really matter about the absolute level of coverage – it would be the same for all. However, this assumes either that the coverage rates are the same for all subject areas, or that all departments will have the same mix of subject areas so that differences would not matter. We can throw some light on this by considering the extent to which these general results encompass more specific variations.

Table 3 looks at the different fields or subject areas covered by the journal papers only. This is very important if the Leiden methodology is used as it normalises citations per paper to the mean for the appropriate field but how does one determine how many fields there should be and what they are? In Table 3 we have taken all the papers and classified them into a field based on the definitions and journals from WoS. We have included in this journals that are not themselves included in WoS. We have then amalgamated 62 sub-categories into 9 major ones.

Generally, the cpp for WoS is under half that of GS but there is quite a degree of variability. Clearly in some instances there are small sample numbers. For the general management field the WoS cpp is 6.3 which is 40% of the GS figure, a ratio that is in general agreement with many of the other evaluations in the literature. It is noticeable that OR/management science has a higher cpp perhaps reflecting its science orientation, and IS and computing has a particularly high GS cpp but this may just be a peculiarity of this sample. Business is particularly low in WoS but in their categorisation business includes finance and it is the case that a particularly high proportion of finance (and accounting) journals are not included in WoS

4 CONCLUSIONS

The knowledge produced by academic researchers is increasingly being judged not just in terms of where it is published but in terms of what impact it is

having. Currently, the major metric for impact is the number of citations that papers, authors, departments or journals receive. This, however, depends on the source from which the citations are counted. The traditional citation index – the Web of Science – is reasonable in the sciences but has poor coverage of social science. In this paper we have compared WoS with a more recent, and rather different, competitor – Google Scholar – on the publications for a university business school. The results show that WoS picks up less than half of the journals, papers and citations found by GS. Moreover, the results differ significantly between subject areas within business and management making it difficult to compare departments or individuals that might have different subject mixes.

Google Scholar, on the other hand, suffers from unreliable data and a lack of transparency about its sources but overall it provides a more comprehensive and less subject-dependent citation resource.

REFERENCES

Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. [Open Access]. *Biomedical Digital Libraries, 3*(7).

Bar-Ilan, J. (2008). Which h-index? - A comparison of WoS, Scopus and Google Scholar. *Scientometrics, 74*(2), 257-271.

Evidence Ltd. (2004). *Bibliometric profiles for selected Units of Assessment*. Leeds: Evidence Ltd.

Harzing, A.-W., & van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics, 8*, 61-73.

HEFCE. (2008a). *Counting what is measured or measuring what counts* (No. 2008/14): HEFCE.

HEFCE. (2008b). *Survey of institutions interested in participating in the pilot of the bibliometrics indicator*: HEFCE.

Jacso, P. (2005). 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.

Jacso, P. (2008). Savvy searching: Google Scholar revisited. *Online Information Review, 32*(1), 102-114.

Mahdi, S., D'Este, P., & Neely, A. (2008). *Citation counts: Are they good predictors of RAE scores?* London: AIM Research.

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

Moed, H., & Visser, M. (2008). *Appraisal of Citation Data Sources*. Leiden: Centre for Science and Technology Studies, Leiden University.

Moed, H., Visser, M., & Buter, R. (2008). *Development of bibliometric indicators of research quality*: Centre for Science and Technology Studies, Leiden University.

Tenopir, C. (2005). Google in the academic library. *Library Journal, 130*(2), 32.

van Raan, A. (2003). The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technology Assessment - Theory and Practice, 1*(12), 20-29.

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