

RQF Pilot Study Project – History and Political Science

Methodology for Citation Analysis

Linda Butler, REPP

November 2006

In many HASS disciplines, standard bibliometric measures have no place in research assessment as the coverage of output achieved by ISI's Web of Science (on which these measures rest) is minimal. However, research undertaken in the Research Evaluation and Policy Project (REPP) has shown that the use of **novel approaches to bibliometrics** may make it possible to extend citation analysis to some of the HASS disciplines. **This pilot study seeks to determine whether there is any potential for using such data in the context of the proposed Australian RQF (or indeed any system-wide research assessment process).** Two representative disciplines have been chosen to test the efficacy of the proposed methodology – political science (social sciences) and history (humanities).

1. Using bibliometrics

This pilot study makes no assumption that judgements about 'quality' can be based solely on citation analysis. In fact, we take exception to any proposal to base assessment of research purely on quantitative data. Bibliometric data answer no single evaluative question in their own right. This information must be seen alongside other measures of esteem, performance, visibility and the testimony of expert peers in the activity being analysed. It is best used in conjunction with a peer evaluation process. The phrase Paul Bourke coined for the correct interaction between bibliometric analysis and peer review is that bibliometric data should provide "a trigger to the recognition of anomaly". Where the two methods do not result in consistent views, the reasons for the differences should be investigated to determine whether they result from problems with the numbers, or whether they highlight information unknown to peers in the field.

2. Limitations of standard measures

Standard citation analyses utilise the Thomson Scientific citation indexes. Originally established by the Institute for Scientific Information (ISI), and still commonly referred to by this name, such analyses rest on the references given by ISI-indexed journals to other publications in the same set of journals. For most disciplines in the humanities and social sciences, this results in a very limited coverage of output for two reasons: much of their output appears in other media, such as books, book chapters, and non-ISI journals; and, ISI's coverage of the disciplines' journals is not as comprehensive, particularly for Australian research, as it is in the experimental sciences.

Table 1, drawn from data provided by universities for an ARC Linkage project undertaken by REPP, demonstrates the coverage typically found in the ISI indexes for all fields of research¹. It is based on the data collected by universities for their annual returns to the Department of Education, Science and Training (DEST).

¹ A full description of the project can be found in Butler L. and Visser M.S., 2006. "Extending Citation Analysis to Non-Source Item", *Scientometrics*, 66(2): 327-343.

Table 1: Distribution of publication output by field, Australian universities, 1999-2001

Field		DEST Publication Categories				ISI Percentage of:	
		Books	Book Chapters	Journal Articles	Conf. Papers	All Pubs	Journal Articles
Band 1	Chemical Sciences	0.2	2.1	95.7	1.9	84.6	88.0
	Biological Sciences	0.3	6.3	90.7	2.7	75.6	81.7
	Physical Sciences	0.1	2.6	90.0	7.3	74.3	82.0
	Medical & Health Sci	0.3	6.3	90.5	2.9	69.3	73.7
Band 2	Agriculture	0.4	5.9	79.0	14.7	63.6	78.7
	Earth Sciences	0.9	7.7	82.2	9.2	60.3	72.7
	Mathematical Sciences	0.7	4.3	83.8	11.2	56.8	67.2
	Psychology	1.5	17.4	76.2	4.9	53.6	69.4
Band 3	Engineering	0.4	2.5	52.0	45.1	37.2	71.0
	Philosophy	6.0	23.8	64.8	5.4	28.1	40.3
	Economics	2.9	24.5	64.5	8.0	24.4	37.2
Band 4	Human Society	3.5	27.8	63.0	5.6	18.7	28.3
	Politics and Policy	5.8	37.3	46.1	10.8	16.5	33.6
	Computing	0.4	4.6	32.8	62.3	15.9	47.8
	History	11.6	34.0	50.6	3.8	14.5	27.6
	Management	1.3	11.7	52.9	34.0	12.6	23.2
	Language	6.5	34.0	51.8	7.6	11.4	19.3
	Education	2.5	19.3	54.5	23.6	9.7	17.2
	The Arts	4.4	20.8	54.5	20.3	9.5	16.0
	Architecture	3.0	17.8	35.6	43.6	6.4	17.7
	Law	4.1	22.1	71.9	1.9	5.4	6.6
Journalism, library	3.4	15.2	57.2	24.2	4.4	7.6	

The output from university departments can be classified into four bands. At the highest level, we identified four fields where ISI journals carried at least two-thirds of the total output reported to DEST – the chemical, biological, physical, and medical and health sciences. In these fields, bibliometric indicators are normally robust. In Band 2 fields, ISI journals account for over half the total output, and standard bibliometric analysis has something useful to say about performance, though some caution needs to be employed in interpreting such data as there is a significant proportion of work not covered. In Band 3 disciplines, ISI journals cover only one-quarter to one-third of the total output and, while there standard bibliometric measures may provide some useful information, to rely solely on such indicators can be very misleading. In Band 4 fields, the use of standard bibliometric measures can not be supported at all – less than one-fifth of their output is in ISI journals.

It is for this reason that these two pilot studies seek to determine whether alternative novel bibliometric measures, based on a wider coverage of output, may have value in a research assessment exercise.

3. Mining the Web of Science for non-source citations

In an attempt to overcome the limitations of standard bibliometric measures, REPP has undertaken extensive research in the ‘mining’ the Web of Science for citations to ‘non-source’ items (i.e. publications appearing in media not indexed by ISI. These may be books, book chapters, conference papers, articles in non-ISI journals, or any other type of publication.). While standard bibliometric measures only utilise the citations to publications in ISI-indexed journals, the database records many additional references to books, book chapters, and articles in non-ISI journals. The two different reference universes are perhaps best described diagrammatically:

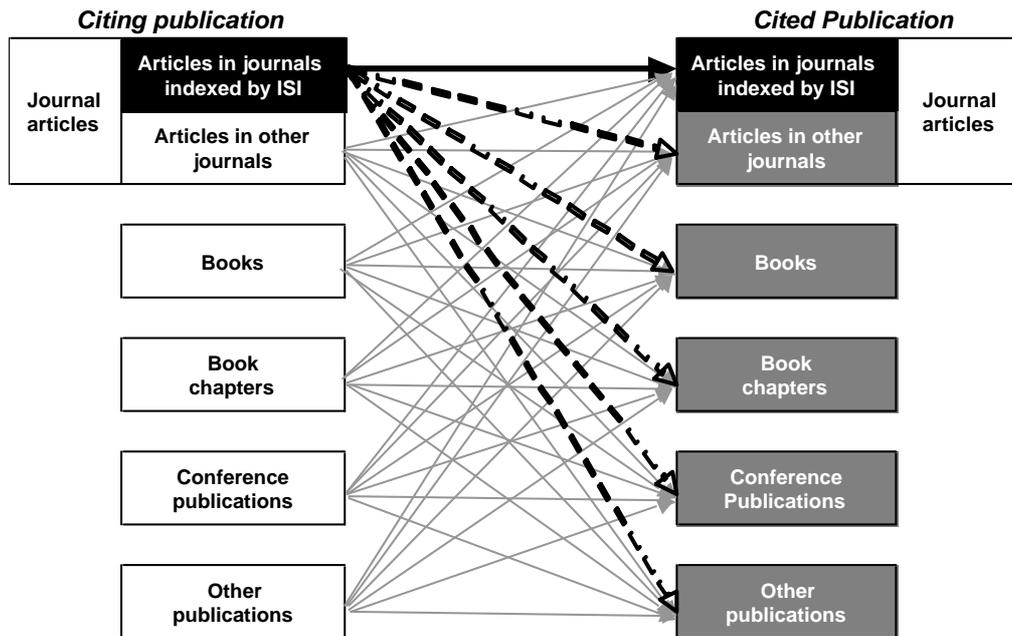


Figure 1: ISI coverage of publications and citations

The references indicated by the bold solid lines are those used in standard citation analysis. We have added the references indicated by the broken bold lines for the analysis presented in this study. As can be seen, it still does not cover references indicated by the grey lines, such as book to book citations.

Citations to non-source items were identified using ISI’s *Cited Reference* search facility, rather than the more commonly used *General Search* query page. A search using the Cited Reference query form returns all cited publications that meet the specified search criteria, whether they are in the indexed journals or in other types of output (i.e. those depicted in the figure above by all the black lines, solid and broken). Using the General Search query form would only have returned citation data on articles in the indexed journals (i.e. the solid black line in figure 1)

Extracting the non-source citations is a time-consuming process, as the references are not standardised in the way citations to ISI journals are, and only the first author of any publication is indexed. This necessitates access to full bibliographic details from either CVs or, as in this case, a list of publications supplied by institutions. Nevertheless, the methodology enables analysts to extract a body of data far greater than that resting solely on ISI journal publications, and for some disciplines this may enable a reasonably robust analysis to be undertaken.

4. The Pilot Study Analysis

19 universities provided details of their DEST publications (excluding conferences) for political science and history. We requested data for the relevant departments, irrespective of the field in which individual academics were publishing. Most data was provided on this basis, though a small number of universities sent details of all publications coded to political science and history, with no information on the academic unit(s) they came from. The data covers a six year period: 2000-2005. The period was chosen to replicate closely the likely length of time covered by the RQF (though not necessarily the exact period) and the expected citation window (i.e. the time frame in which publications could attract citations).

The citation data for all publications was extracted from ISI's Web of Science in October , and multiple instances of varying references to the same publication were aggregated to provide 'clean' data. The data was then aggregated by discipline and institution, and a range of measures calculated:

- total citations – for all publications, and for each type of publication (books, book chapters, and journal articles);
- citations per publication – the total number of citations was divided by the number of publications reported by the institution. Calculations were undertaken for all publications, and for each type of publication; and
- ISI citation rates – citation per publication calculations were made for journal articles, limited to those in journals indexed by ISI (i.e. the equivalent of a standard bibliometric measure).

5. Pilot Study Questions

A set of tables will be provided for each discipline. Initial tables will be distributed with this overview of the methodology, prior to the workshop. Other detailed analyses may also be presented on the day. We are keen to seek your assessment of the measures and the data on which they are based.

The most important questions we will be seeking to answer at the workshop are:

- Does the 'picture' painted by the data coincide with your knowledge of the relative strengths and weaknesses of the discipline in the participating universities?
- Where the data appear at odds with your knowledge of the discipline – are there any factors that immediately spring to mind that might cause this?
- Where the data reinforce your assessment – which measures are the most robust?
- Does the data provide additional information that would assist RQF panellists in assessing the field?