Can the Impact of non-Western Academic Books be Measured? An investigation of Google Books and Google Scholar for Malaysia[1]

Abrizah Abdullah
Department of Library & Information Science, Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia
Mike Thelwall
Statistical Cybermetrics Research Group, University of Wolverhampton, Wolverhampton, UK.

Citation indicators are increasingly used in book-based disciplines to support peer-review in the evaluation of authors and to gauge the prestige of publishers. However, since global citation databases seem to offer weak coverage of books outside the West, it is not clear whether the influence of non-Western books can be assessed with citations. To investigate this, citations were extracted from Google Books and Google Scholar to 1357 Arts, Humanities and Social Sciences (AHSS) books published by five university presses during 1961-2012 in one non-Western nation, Malaysia. A significant minority of the books (23% in Google Books and 37% in Google Scholar, 45% in total) had been cited, with a higher proportion cited if they were older or in English. The combination of Google Books and Google Scholar is therefore recommended, with some provisos, for non-Western countries seeking to differentiate between books with some impact and books with no impact, to identify the highly cited works or to develop an indicator of academic publisher prestige.

Introduction

Books published in the arts and humanities are "the main course of a meal" (Williams et al., 2009, p.76), especially if they are published by prestigious university presses and widely sold, read, reviewed, cited, and purchased by libraries (Gump, 2006; Zuccala & van Leeuwen, 2011). For research assessors seeking to evaluate the contribution of books without reading them, the prestige of a press has perhaps been the most obvious characteristic to check because statistics about the other facets may be difficult to calculate. Moreover, whilst citation counts have been routinely used to estimate the impact of journal articles for a long time, the use of citations to estimate the impact of books is a more recent phenomenon, presumably because of the lack of major book citation indexes. The Web of Science (WoS), Scopus and Google Scholar now all include citations *to* books (e.g., Bar-Ilan, 2010; Butler & Visser, 2006), as do some discipline-specific databases such as PsycINFO and Sociological Abstracts. In WoS, citations to books from source items are in the "cited reference search" tab, and in Scopus these are available by clicking on the "view secondary documents". However, none of these databases seem to have extensive coverage of citations *from* books and all seem to be biased towards publications from Western nations (Dassa, Kosmopoulos & Pumain, 2010; Shelton, Foland & Gorelsky, 2009). Hence it is unclear whether it is currently possible to use currently available tools to assess the impact of books published outside of the West. This is important not only for the non-Western countries involved but also for science as a whole due to the increasing academic contributions of countries outside the West, and Asia in particular (e.g., Leydesdorff & Wagner, 2008; Zhou & Leydesdorff, 2006).

In response to the need for book impact assessment, the Book Citation Index (BKCI) in WoS was made available in October 2010, indexing over 30,000 editorially selected books from 2005 onwards to make bibliometric analyses of Arts, Humanities and Social Sciences (AHSS) books possible. Priority was given to books and book series with higher citation impact. This database has been used to evaluate the impact of individual books (Torres-Salinas, Rodríguez-Sánchez, Robinson-Garcia, Fdez-Valdivia & Garcia, 2013) and the

---

prestige of publishers (Torres-Salinas, Robinson-Garcia, Jimenez-Contreras & Lopez-Corza, 2012; Torres-Salinas, Rodríguez-Sánchez, Robinson-Garcia, Fdez-Valdivia & Garcia, 2013).About 75% of the book chapters indexed in BKCI are from the USA and England (Torres-Salinas, Rodríguez-Sánchez, Robinson-Garcia, Fdez-Valdivia & Garcia, 2013), however, casting doubt on whether it is useful for citation analysis outside of these countries. In 2013, Scopus began a 3-year initiative to increase its coverage beyond the current 360 book series to 75,000 titles. This expansion will include monographs, edited volumes, major reference works and graduate level textbooks, aimed at meeting the needs of book-oriented disciplines in the AHSS.

Google Books and Google Scholar are alternative sources for book citation counts. Although Google Books does not report book citations, it supports full text searching of its database so that potential citations to a book can be identified by searching for the book title and eliminating false matches (Kousha & Thelwall, 2009). Google Scholar has also been reported to have substantial numbers of citations *to* academic books (Kousha, Thelwall, & Rezaie, 2011). Because of the internationally limited coverage (so far) of BKCI, Google Books and Google Scholar seem to be the logical candidates for impact assessment of non-Western books. Nevertheless, Google does not release information about which books it has indexed and which libraries it has visited to scan books, and so it is not clear whether the coverage of Google Books or Google Scholar are sufficient for impact estimation of books published outside of the West. Moreover, no study has investigated this issue before and so, to partially fill in this gap, this paper examines whether Google Books and Google Scholar can provide adequate citation evidence for books published in a non-Western country.

Literature Review

Academic books typically collate experiences, document observations, preserve knowledge for future generations, and communicate findings to academics, students and practitioners (Serenko, Bontis & Moshonsky, 2012). The quality of scholarly books has risen to the current situation in which they are typically considered to be a source of authority (Saenger, 1975). The scholarly book is also an essential part of promotion and tenure in major US universities in some disciplines (Cronin & La Barre, 2004). To emphasise their importance, Adler and Harzing (2009) claim scholarly books receive more than 40% of all citations, and are also often read by non-academics and practitioners.

Since books are important in academia, methods are needed to assess their impact or value. Torres-Salinas and Moed (2009) suggested using library catalogues, such as WorldCat, as a data source for book impact by counting the number of libraries holding each book. Similarly, White et al. (2009) introduced the term "libcitation", which, for a particular book, "increases by 1 every time a different library reports acquiring that book in a national or an international union catalogue" (p. 1084). The libcitation idea was a partial response to the inadequacy of metrics in Australia for assessing AHSS scholars. Since exploratory trials found no correlations between citation and libcitation counts, however, they seem to be measuring different things (White et al., 2009) and it is possible that libcitations assess the wider readership of a book, as estimated by librarians, rather than its academic impact.

The assessment of AHSS books through citations is an established theme within bibliometrics (Cronin, Snyder & Atkins, 1997; Cullars, 1992; Lindholm-Romantschuk, 1996; Nederhof, 2006; Stone, 1982; Thompson, 2002; Torres-Salinas, Rodríguez-Sánchez, Robinson-Garcia, Fdez-Valdivia & Garcia,2013).For example, citation analysis has been used to measure the impact of open access (OA) publishing of academic books and Google Books (Snijder, 2010); OA publishing enhanced discovery, but did not affect sales figures or citation rates. On a much smaller scale, Hammarfelt (2011) analysed the citations (more than 4,000 in WoS) to a single book, Walter Benjamin's *Illuminations* (1968/2007),combining them with qualitative investigations to map the impact and dissemination of this book. Downloads for digital books are a possible future impact indicator, although download data might not be accessible for large scale studies from some academic publishers (Kousha, Thelwall & Rezaie, 2011). Previous studies have suggested that Google Scholar (Bar-Ilan,

2010; Garcia-Perez, 2010; Mingers & Liptakis, 2010; Kousha, Thelwall & Rezaie, 2010; 2011) and Google Books (Kousha, Thelwall & Rezaie, 2011) contain publication types outside of WoS and Scopus and therefore might be particularly useful for book impact assessment in AHSS. When studying citations to the book "Introduction to informetrics" in Google Scholar, WoS and Scopus, Bar-Ilan (2010) concluded that the databases supplemented each other and had comparable performances. When limiting the publication year to 1996 onwards, Google Scholar's coverage was "surprisingly good, and its accuracy was also better than expected" (pp. 505).Studies by Garcia-Perez (2010) for psychology publications and Mingers and Liptakis (2010) for business and management publications confirmed that Google Scholar reported substantially higher publication and citation metrics than did WoS. Kousha, Thelwall and Rezaie (2011) examined citation counts to 1,000 books submitted in the 2008 UK Research Assessment Exercise and found citation counts in Google Scholar to be more than three times as high as those from in Scopus. The median of the Google Books citations was more than double the median of the Scopus citations, suggesting that Google Books could be a valuable source of citation impact for book-based disciplines. In conclusion, both Google Books and Google Scholar seem to be promising for evaluating book-based disciplines, but Google Books is more practical for large-scale analyses because it allows automated searching.

Many academic books are published by university presses, some of which have specialized areas of expertise, and these seem to be a particularly important source (Thompson, 2002). Academics in book-based disciplines, such as in the arts, humanities and many social sciences, may seek to publish their books with the most prestigious presses in their field. Librarians also use the reputation of the publisher as a prominent criterion in the selection of books (Jordy, McGarth & Rutledge, 1999) and it is also used for research assessment purposes in the arts and humanities where books are most important (Harzing, 2010). They may also prefer to publish outside their own university press to avoid any suggestions of favouritism and to provide external validation (Brown, Griffiths & Rascoff, 2007). Since press prestige is sometimes used to evaluate books, methods are also needed to systematically evaluate publication houses. For example, one study surveyed academic librarians for their opinions of publishers of books related to education, producing a list of 40 high reputation publishers (Sieber & Gradmann, 2011). In contrast Torres-Salinas et al. (2013) used a citation-based metric with BKCI data to identify the 20 most productive publishers in four disciplines: Sciences, Engineering & Technology, Arts & Humanities and Social Sciences. Another study assessed the proportions of books from a publisher that were indexed in various book databases (Tausch, 2011).

In summary, whilst many methods, including citation-based metrics from various sources, have been used to evaluate books and publishers, there is no one accepted method and source of data. Moreover, there is no evidence to suggest that the international coverage of databases that include book citations is adequate for assessing books published in non-Western countries. Google Books seems to be a particularly promising source of citations in this regard because it seems to be the largest source of full text books. Even though it is probably biased towards the US and the West in general, university libraries in the West may have copies of overseas books for their international studies sections and these may thus find their way into Google Books via its university library scanning programmes. Google Scholar is another promising source of citations, especially if more of the national academic output is published online, because it encompasses a wide range of non-traditional academic sources from the Web.

Research Questions
The main objective of the study is to investigate whether Google Books and Google Scholar can provide good impact data for books published in non-Western countries, given that their publications may not be well covered by major international citation databases such as the Web of Science and Scopus. The scope is restricted to arts, humanities and social sciences books because these are the areas in which books are often key scholarly outputs (Williams et al., 2009; Gump, 2006; Zuccala & van Leeuwen, 2011). The empirical part of this study

focuses on one non-Western country, Malaysia, which has a developed and well-defined local academic publishing industry based in its universities. Although the intention of the research is to show that Google Books is internationally useful, including in non-Western countries, the focus is on a single non-Western country. The purpose is therefore to show that it is *possible* for Google Books to useful for a non-Western country rather than to show that Google Books is useful for *all* non-Western countries. Each country has its own publishing industry, politics and linguistic traditions and so it is likely that Google Books will not be useful for some countries and that the extent to which it is useful will vary by country. Nevertheless, demonstrating that it can be useful in a non-Western, non-English context can pave the way for any country to investigate its use in their science evaluation system. The following questions drive this research:

1. Are there sufficient citations to assess the impact of academic books published in Malaysia using Google Books and/or Google Scholar?
2. Does the answer to the above question vary by publication type, publication language, publication year and broad discipline (social science vs. arts & humanities)?
3. Is it possible to identify high and low impact Malaysian university publishers based upon Google Books and/or Google Scholar citations to their books?
4. Which types of Malaysian books achieve a high citation impact?

Method

A comprehensive list of AHSS books was obtained from the five main university presses in Malaysia: University of Malaya (UM); University Kebangsaan Malaysia (UKM); Universiti Putra Malaysia (UPM); Universiti Sains Malaysia (USM) and Universiti Teknologi Malaysia (UTM). These universities have "research-intensive university" status and are ranked in the top 100 of the 2013 QS University Rankings Asia. The QS rankings are flawed and are only approximate indicators of research quality since they combine prestige and citation data with teaching and financial data (Aguillo, Bar-Ilan, Levene, & Ortega, 2000). The QS rankings can also be affected by errors for individual universities, including in Malaysia (Holmes, 2006). Nevertheless, the QS have a medium level of similarity with two other (Asian) ranking schemes for the top 500 world universities (Aguillo, Bar-Ilan, Levene, & Ortega, 2000) and so, in the absence of a clearly best candidate seem to be a reasonable choice as an approximate indicator of research prestige for Malaysian universities. Research universities in Malaysia consider books published by university presses as a prerequisite for AHSS faculty promotion and tenure and so they are an important publishing outlet for the country's research. The five lists were compared with a list of AHSS books published by the Malaysian Scholarly Publishing Council (MAPIM) for completeness and no gaps were found. Edited books were included in the lists since the focus of the study is on the publishers and the books they publish, and not on the authors. We removed edition information from the list because citations to any edition of a book can reflect its intellectual impact (Kousha &Thelwall, 2009). After removing duplicates (e.g., multiple editions of the same book) the final list contained 1,357 books, 967 in Malay and 390 in English. This is unsurprising given that AHSS scholars often publish in their native languages (Rauvargers, 2011). The number of books published per year has risen steadily since the first was published in 1961 (Figure 1) so the results of this paper refer mainly to books published since the year 2000. The drop is perhaps due to a recent of Malaysian research-based university emphasis on publishing in scholarly journals indexed by major citation databases such as WoS and Scopus. The overall profile of publishing shown in Figure 1 is important to bear in mind when interpreting the results.
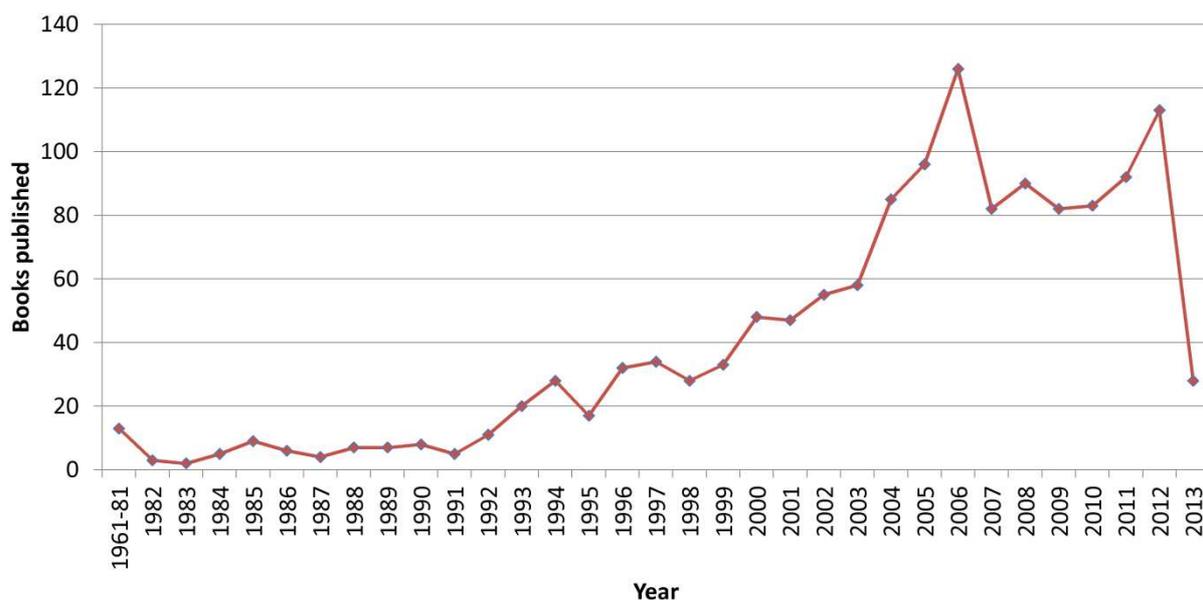
Figure 1. The number of Malaysian university press books published each year.

The final list included publisher subject categories (anthropology, art, business & management; communication & media studies; economics; education; geography, history; humanities, language & literature; law; psychology, religion and sociology) and these were re-categorized into two, Arts & Humanities and Social Sciences, based on the Web of Science subject categories.

Webometric Analyst[2] was used to search for citations from Google Books via its Applications Programming Interface. For each book, the query used comprised the book titles in quotes (a maximum of the first 4 words), together with the list of the authors' last names followed by their initial(s) and the year of publication (e.g. 1966). We used the standard citation style used for Chinese, Indian and Western names; for Malay names, the father's name is normally the last name, whereas the personal and second names or other names that follow are treated as initials. For example, Wan Ahmad Tajuddin bin Wan Abdullah is treated as Abdullah, WATW (the bin or binti, referring to son or daughter of, is not included). Malaysian names are often cited by first name rather than last name, (e.g., Abrizah A. rather than A. Abdullah for Dr Abrizah binti Abdullah) but testing suggested that the vast majority of results citations would be successfully retrieved by using only the last name. Google Books returns approximate matches rather than exact matches and so Webometric Analyst was used to automatically filter out false matches by searching the description field of each result for a mention of the book title and author, and also by removing standard text indicating an advert within a book or a book catalogue. The remaining citations were manually checked for any remaining false matches and to ensure that a book had been mentioned as a citation rather than in an advert, author introduction, or annotated bibliography. A total of 66 non-citations (from a total of 2254) were manually removed in the final stage, so the accuracy rate for Webometric Analyst was 97.1%.

For Google Scholar searches, we manually searched the exact titles of books as phrase searches and combined them the first author surnames, with other bibliographic data such as publishers or publication years if necessary. We then recorded the number of Google Scholar citations identified by the 'cited by' option (e.g., cited by 28) below each displayed record of the exact match. For Malay names (and in some cases Chinese names) we had to double check the citations under variant names of the cited authors, since for some Malaysian publications, the Malaysian convention may be used and the first names

---

[2]Available free at: http://lexiurl.wlv.ac.uk.

are referred to in full in the text and reference list. The citation counts reported by Google Scholar may include duplicate citing items or false matches. For this reason we manually checked the full text of citing documents through either the preprint/postprint links (e.g., "[PDF] from academia.edu") from open access repositories or we followed the link in Google Scholar to the full-text of the citing documents through institutional subscriptions to major journal publishers. In some cases, especially for edited books, it was not possible to check the citations in the context of the retrieved documents through these methods. Therefore, the only practical method was to recognise the formal citation based on the brief textual information below each retrieved record in the Google Scholar search results. Our observation was consistent with Kousha and Thelwall (2009) in that if the citation information of a book appeared in brief records in bold and looked like a cited reference (e.g., APA or Chicago citation styles), then it was likely to be a formal citation, otherwise it was more likely to be a false match.

All Google Books and Google Scholar searches were conducted during June 2013 for each book to lessen the potential impact of the time window on the citation count. However this short period of time may still have a small impact on the volatile citation changes over time, especially in Google Scholar, for the subject disciplines of the sampled books.

The data was mainly analysed with differences in proportions tests to compare the proportions of cited and uncited books between various subgroups in the data. It would also have been possible to compare the total number of citations between subgroups using a Kruskal-Wallis non-parametric ANOVA but the difference in proportions test is preferable because it is not sensitive to unusual citation counts for individual books.

RESULTS

Research Question 1 (Impact of Malaysian Books)
About 45% of the Malaysian academic books had a measurable impact in either Google Books or Google Scholar (Table 1). Hence, the combination of the two would be sufficient for research assessment that had the goal of differentiating between higher and lower impact work in such a way that fine-grained impact assessment for the bottom half was not needed. Individually, these sources would be sufficient if the bottom two thirds (Google Scholar) or three quarters (Google Books) did not need fine-grained differentiation.

The overlap between Google Books or Google Scholar was quite small, with (15%) had impact in both and 45% having an impact in at least one. Google Scholar contained citations to more Malaysian academic books (37%) than did Google Books (23%) but the low overlap suggests that it is best to combine both sources for impact evaluation purposes.

Table 1. Citations to Malaysian books by subject category, using Google Books (GB) and Google Scholar (GS).

Subject Category	Books	Books with citations in GB	Citations per book in GB	Books with citations in GS	Citations per book in GS
Art	36	3 (8%)	0.64	9 (25%)	1.92
History	84	27 (32%)	1.96	41 (49%)	3.46
Humanities, multi-disciplinary	46	8 (17%)	1.74	14 (30%)	1.13
Language & Literature	250	50 (20%)	2.08	98 (39%)	1.70
Law	46	14 (30%)	1.93	19 (41%)	1.76
Religion	148	18 (12%)	0.80	61 (41%)	1.71
Arts & Humanities	610	120 (20%)	1.63	242 (40%)	1.92
Anthropology	4	1 (25%)	0.75	2 (50%)	0.75
Business & Management	114	26 (23%)	0.54	29 (25%)	0.67
Communications & media studies	59	18 (31%)	1.42	19 (32%)	2.08
Economics	157	37(24%)	1.40	51 (32%)	1.43
Education	238	64 (27%)	1.31	87 (37%)	1.47
Geography	15	3 (20%)	1.93	6 (40%)	2.53
Psychology	4	1 (25%)	0.25	1 (25%)	0.50
Sociology	156	44 (28%)	3.01	62 (40%)	1.54
Social sciences	747	194 (26%)	1.58	257 (34%)	1.42
Total	1357	314 (23%)		499 (37%)	

As could be expected, the citation impact of Malaysian books is highly skewed. In Google Books, a total of 103 books were cited only once; and only 50 books were cited more than 10 times. In Google Scholar, 192 books were cited only once, and 51 books were cited more than 10 times.

Research Question 2 (Variations of Impact by language, publication type, broad discipline and year)

Almost three quarters (71%) of the books were in Malay and the rest were in English. The English books were significantly more likely to be cited in both Google Scholar (48% vs. 32%, differences in proportions test z=5.30, p=0.000) and Google Books (40% vs. 16% in Google Books, differences in proportions test z=9.35, p=0.000).

Only 7% of the books were edited and there was no significant difference in the extent to which they were cited in Google Books (20% vs. 23%, differences in proportions test z=0.82, p>0.05) but edited books were cited more often in Google Scholar (53% vs. 36%, differences in proportions test z=3.22, p<0.01). The greater citation of edited works in Google Scholar makes sense in that book chapters may be more common in the social sciences and in fields that combine journal articles and edited books.

Almost half (45%) of the books were from Arts & Humanities and these were less likely to be cited in Google Books (26% vs. 20%, differences in proportions test z=2.74, p<0.01), although the differences are not large. Surprisingly, Arts & Humanities books more likely to be cited in Google Scholar than Social Sciences books (40% vs. 34%, differences in proportions test z=2.00, p<0.05).

The proportion of books cited in Google Scholar stayed very approximately constant until 2005 (the greater variability being a consequence of the lower numbers of books involved) and then decreased in two steps (Figure 2). Since the citation searches were conducted in 2013, this suggests that a gap of at least 2 years is needed to obtain a minimal number of citations but a gap of 8 years from publication is required to be reasonably certain of being able to separate out the cited from the uncited books, at least in Google Scholar. The situation for Google Books essentially the same, but with lower numbers and with citations from books perhaps steadily increasing over a longer period of time.
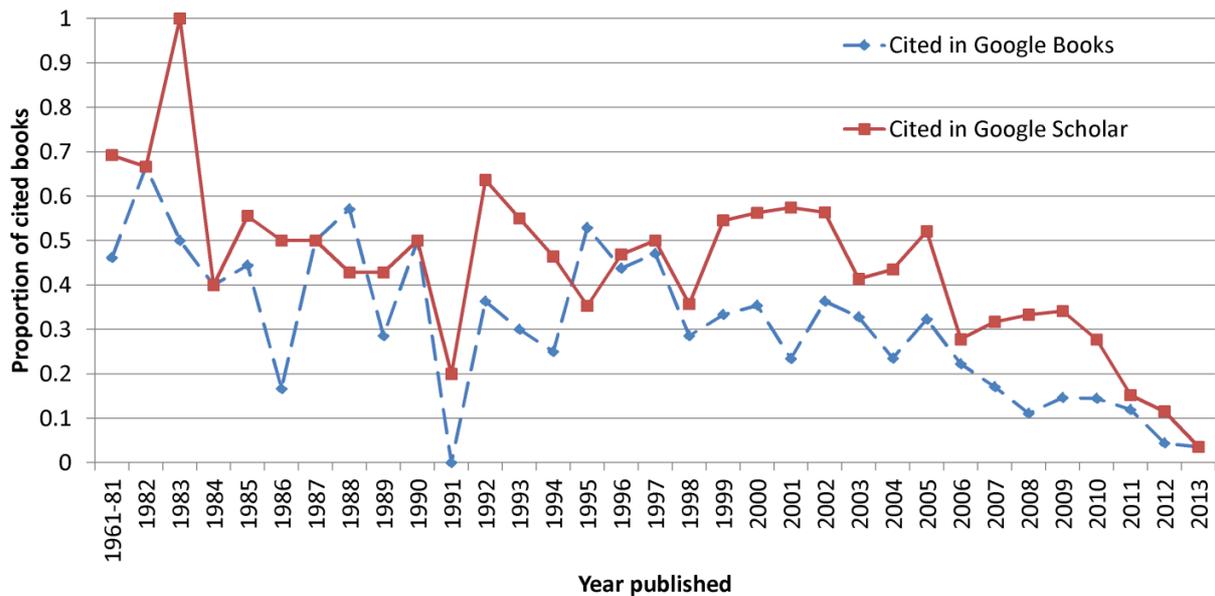
Figure 2. The proportion of Malaysian university press books cited by publication year.

Research Question 3 (Comparison of Malaysian University Presses)

The university presses were compared on the basis of the proportion of cited books because, as discussed above, this is a more robust test than one relying on citation counts. Hence the underlying assumption is that better university presses will tend to publish a higher proportion of cited books. The five Malaysian university presses have all published substantial numbers of books (Table 2) but do not have the same proportion of books cited in Google Scholar (chi-square 25, p=0.000) and do not have the same proportion of books cited in Google Books (chi-square 92, p=0.000). The differences qualitatively seem large enough to differentiate between higher and lower impact presses, at least in theory. Nevertheless, the reason for the difference is not clear and it does not seem to associate with the quality of the university involved. For example the lowest ranked university, according to QS University Rankings Asia, has the second largest proportion of cited books. This could be for reasons unrelated to book quality, however, such as a particularly successful marketing team.

Table 2. The proportion of cited books from five Malaysian university presses, using Google Books (GB) and Google Scholar (GS).

Publisher	Books	Books with GB citations	Books with GS citations	QS University Rankings Asia
UM Press	358	38%	44%	33
UKM	458	18%	38%	57
USM	229	14%	25%	61
UTM	181	10%	31%	68
UPM	131	36%	40%	72
Total	1357	23%	37%	

Research Question 4 (Characteristics of Most Frequently Cited Books)

The most highly cited books in both Google Books and Google Scholar were examined (tables 3 and 4). Eight books were in the top 20 highly cited in both databases. The following characteristics were found that set them apart from the remaining books. All the differences reported are statistically significant with p<0.01 using a differences in proportions test except the single case noted.

a) They are mostly written in English (95% in Google Books; 85% in Google Scholar; only 29% are English in the initial list).
b) They are disproportionately published by UM Press (60% in Google Books; 35% in Google Scholar; 26% are from UM Press in the initial list).
c) They are disproportionately single-authored (80% in Google Books (not significant); 95% in Google Scholar; 66% are single-authored in the initial list).

As listed above, the most striking difference is that English books are far more likely to be highly cited than books in Malay. Although this may partly be explained by some of the books being written by scholars visiting Malaysia from prestigious Western universities (e.g., James C. Scot), the highly cited authors predominantly have names that do not have a Western origin (16 out of 20 in Google Books; the exceptions are 1, 3, 5 and 19 in Table 3; 18 out of 20 in Google Scholar; the exceptions are 1 and 20 in Table 4). Several explanations are possible for this. Publishing in English could be an advantage for attracting citations or for citation coverage in both databases, better quality work may tend to be published in English, or disciplines tending to publish in English may tend to be more highly cited (e.g., more quantitative).

Since UM Press has the highest proportion of cited works in both databases and a disproportionate share of the highly cited works and UM is the highest ranked university (Table 2) it seems reasonable to hypothesise that UM Press tends to publish the best books. Single authored books are statistically significantly overrepresented in the Google Scholar top 20 list but it is not clear why. Solo authored articles tend to be less cited than multiple authored articles in science (e.g., Costas, & van Bochove, 2012) and the same has been shown in some social sciences, with national variations (e.g., economics: Levitt & Thelwall 2010) but there seem to be no research on any links between collaboration and citation for books in any area or for arts and humanities research. A possible explanation is a culture for leading arts, humanities, and/or social science scholars in Malaysia to not collaborate (see below).

The most highly cited book in Google Books was a joint publication between UM Press and Yale University with the same ISBN. According to the cited reference search in WoS, the 1985 edition of this book has been cited 237 times in articles, book reviews and books (recall that Google Books only counts book citations). In the case of this influential book, WoS reports a substantially higher citation count than extracted from Google Books. Google Scholar combines together citations from different versions and treats them as a single citation (e.g., 6804). However, Google Scholar reported a much lower citation count (20, Table 4) for the 1985 edition, but a much higher value (more than 6000) for the reprints. This book was authored by an influential American professor at the prestigious Yale University, which accounts for its success and it is hence a special case. It is not clear why the book was dual published, but it may be due to the research described having been conducted within Malaysia (from USM) and on Malaysians (peasant resistance strategies in Kedah State), even though the theory expounded was generalised rather than specific to Malaysia.

The most highly cited book in Google Scholar was by a British professor from King's College London who investigated the historical geography of South-East Asia and China, and was appointed to a lectureship at the University of Malaya (Singapore) in 1952. The book was written during the visit, published in 1961 by UM Press (reprint 2010) and distributed by Oxford University Press. This book describes how Buddhist pilgrims, Chinese, Indian, and, later, European explorers slowly discovered the Golden Khersonese, the ancient name for the Malay Peninsula (currently Malaysia). This book was the third book highly cited in Google Books and received 57 citations in WoS.

Some of the highly cited books in Google Books were authored by leading Malaysian scholars in their respective field of studies, such as Naguib Al-Attas (contemporary Muslim philosophy), Osman Bakar (science philosophy), Jomo K.S. (international economics), Khoo Kay Kim (Southeast Asian history) and Noordin Sophiee (international politics). Their books in Table 3 have been cited at least once in WoS. Their other books published with prestigious university presses such as Oxford and Cambridge have been highly cited in WoS

(e.g., Jomo, K.S. with 815 citations). This is some evidence that the Google Books citations are relevant to academic achievements. The fifth book is a little unusual in that it was authored by an American professor from the Smithsonian Institute whilst visiting UKM. The nineteenth book is an edited book also by an American Professor during his six years research career at USM. Apart from Noordin Sophiee, these aforementioned scholars did not appear in the list of top 20 frequently cited books in Google Scholar.

Table 3. The top 20 most frequently cited books published by five Malaysian university presses, using citations extracted from Google Books.

	Book Title/Author/Year	Citations GB (GS)	Subject
1	*Weapons of the weak: Everyday forms of peasant resistance.* James C. Scot. 1985.	298 (20)	SS
2	*English in a globalised environment: Investigating an emerging variety of English.* Zuraidah Mohd Don. 2006	165 (4)	AH
3	*The Golden Khersonese. studies in the historical geography of the Malay Peninsula before AD 1500 (1961).*Paul Wheatley. 1961, 2010 (reprint)	106 (117)	AH
4	*The oldest known Malay manuscript: A sixteenth century Malay translation of the 'Aqa'id of al-Nasafi.* S.M. Naguib Al-Attas. 1988.	64 (5)	AH
5	*Educating in a globalizing world: Why the reach exceeds the grasp* Joan M. Nelson.2007.	49 (3)	SS
6	*Competencies needed by teachers.* Habibah Elias, Zaidatol Akmaliah Lope Pihie&RahilMahyuddin.2005.	48 (1)	SS
7	*Dilemma of independence: Two decades of Malaysia's foreign policy, 1957-1977.* J.Saravanamuttu. 1983	34 (35)	SS
8	*Language testing: The construction and validation.* Hassan Basri Awang Mat Dahan. 2002.	33 (24)	AH
9	*Law and the electoral process in Malaysia.* S. Sothi Rachagan. 1993.	33 (1)	AH
10	*Educations and politics in Indonesia 1945-1965.* Lee Kam Hing. 1985.	32 (21)	SS
11	*Islam and civilization dialogue: The quest for a truly universal civilization.* Osman Bakar. 1997.	26 (14)	AH
12	*Coastal resources of West Sabah: An investigation into the impact of oil spill.* Chua Thia-Eng&J.A.Mathals (ed.), 1978.	26 (12)	SS
13	*International communication: English language challenges for Malaysia.* Saran Kaur Gill. 2002.	26 (41)	AH
14	*Industrial policy in East Asia: Lessons for Malaysia.* Jomo K.S. & Tan KockWah.2000	26 (18)	SS
15	*SejarahmasyarakatMelayumoden.* Khoo Kay Kim.1985. (English translation: *Modern Malay history*)	23 (7)	AH
16	*From Malayan Union to Singapore Separation Political Unification in Malaysia Region 1945-1965.* Mohamad Noordin Sopiee. 2005.	22 (65)	AH
17	*Language Choice & Communication in Malaysian Business.* Shanta Nair-Venugopal. 2000.	22 (30)	SS
18	*Corporate Governence: An Islamic Paradigm.* Vaseehar Hassan, Bala Shanmugam & Vignesen Perumal. 2005.	22 (1)	SS
19	*Farmers in The Hills: Upland Peoples of North Thailand.* Anthony R. Walker (ed.) 1975.	21 (19)	SS
20	*Islam and Confucianism a Civilizational Dialogue. Osman Bakar.* 1975.	21 (12)	AH

Table 4. The top 20 most frequently cited books published by five Malaysian university presses, using citations extracted from Google Scholar.

	Book Title/Author/Year	Citations GS (GB)	Subject
1	*The Golden Khersonese. studies in the historical geography of the Malay Peninsula before AD 1500 (1961).*Paul Wheatley. 1961, 2010 (reprint)	117 (106)	AH
2	*From Malayan Union to Singapore Separation Political Unification in Malaysia Region 1945-1965.* Mohamad Noordin Sopiee. 2005.	65 (22)	AH
3	*Penyelidikan Pendidikan. Mohd Najib Abd Ghafar. 1999.* (English translation: *Educational research*)	54 (0)	SS
4	*Resisting Colonialist Discourse.* Zawiyah Yahya. 2003.	47 (4)	AH
5	*International Communication: English Language Challenges for Malaysia.* Saran Kaur Gill. 2002.	41 (26)	SS
6	*The Pondok & Madrasah in Patani.* Hasan Madmarn. 2002.	37 (0)	AH
7	*Dilemma of independence: Two decades of Malaysia's foreign policy, 1957-1977.* J.Saravanamuttu. 1983	35 (34)	SS
8	*Iban Politics & Economic Development : Their Pattern & Change.* Jayum A Jawan. 1994.	31 (0)	SS
9	*Language Choice & Communication in Malaysian Business.* Shanta Nair-Venugopal. 2000.	30 (22)	SS
10	*Iban Adat and Augury.* Benedict Sandin. 1988.	30 (0)	AH
11	Orang Rejimen : The Malays of The Ceylon Rifle Regiment. B.A Husainmiya. 1990.	30 (0)	AH
12	*Law and the electoral process in Malaysia.* S. Sothi Rachagan. 1993.	24 (33)	AH
13	*A Malaysian English primer: aspects of Malaysian English features.* Loga Mahesan Baskaran. 2006.	24 (0)	AH
14	*Islamic Modernism in Malaysia the Life and Thought of SayidSyekh al-Hadi 1867-1934.* Ibrahim Abu Bakar. 1994.	23 (8)	AH
15	*Stock Pricing in Malaysia, Corporate Financial & Investment management. Mohamed Ariff, Shamsher Mohamad & Annua rMd Nassir.* 1998.	22 (4)	SS
16	*Educations and politics in Indonesia 1945-1965.* Lee Kam Hing. 1985.	21 (32)	SS
17	*Malaysian Industrial Relations: Law and Practice. V. Anantaraman. 1997.*	21 (10)	SS
18	*Reka Bentuk Tinjauan Soal Selidik Pendidikan.* Mohamad Najib Abdul Ghafar. 2003. (English translation: *Educational survey research design*)	21 (0)	SS
19	*Statistik Penyelidikan Dalam Pendidikan Sains Sosial. Alias Baba. 1997.* (English translation: *Educational statistics in social sciences*)	21 (0)	SS
20	*Weapons of the weak: Everyday forms of peasant resistance.* James C. Scot. 1985.	20 (298)	SS

Limitations and discussion

The main limitation of this study is that it only covered one country and the results may be different in other countries. In particular, the results might be worse for countries that publish extensively in languages with more complex characters, such as Chinese or Arabic, which may be more difficult for Google to scan or process (although both Google Books and Google Scholar appear indexed all languages that we checked, so there does not appear to be a simple technical limitation). A second limitation is that the extent to which a nation's academic books are cited probably depends at least partly on its investment in research and education, and hence its wealth (King, 2004). Thus the results for poorer countries may well be not as good. Moreover, since Google is based in the US, its coverage of international books may be skewed towards countries that the US has had an interest in, or which are relevant to the national interest, but Malaysia does not seem to qualify in either case and so it seems unlikely that Malaysian books are covered better by Google than average for its level of wealth. In summary, the results for other countries may be worse than those for Malaysia if those countries are significantly poorer or use a script that is difficult to automatically accurately scan. The results are likely to be better for significantly richer

countries and perhaps also for those that are relevant to US historical or national interests (e.g., South and Central America, the Middle East, Vietnam, North and South Korea, China).

There are technical and coverage differences between Google Books and Google Scholar that are important to give context the results. Google Scholar apparently identifies documents from web crawling and publishes guidelines for authors and publishers to ensure that their documents online are in a format that Google Scholar can properly identify (Google, 2013a). Google (2003b) claims that "Google Scholar generally reflects the state of the web as it is currently visible to our search robots and to the majority of users." Google Scholar probably inherits the international coverage of the main Google search engine (but with some national biases, see: Vaughan, & Thelwall, 2004) that presumably provides the raw data via its crawlers. Google Books does not seem to use web crawling but scans books from specific libraries and accepts books to index from publishers and authors (Google, 2013c). Books that are not in the public domain seem to be still full-text searchable, even if they cannot be fully viewed. Google Books also includes metadata for other books, presumably from large international catalogues, but this metadata probably does not contain many citations so is not relevant here. Google Books aims "to include books from all the world's languages and cultures" and works "with publishers from many countries around the world to include their books in the Partner Programme" (Google, 2013d). Nevertheless, it seems likely that Google Books is US and English-centric due to its scanning policy, which presumably started in the USA, and perhaps also mirroring the international uptake of the Google search engine to some extent for books from publishers and authors.

Given the differences found between the citations in the fields within Arts & Humanities and Social Sciences in this study (Table 1), generalizations to individual AHSS fields could be misleading. Unfortunately, the numbers involved are too small for effective statistical conclusions to be drawn about individual fields and so it is possible that the results do not apply to some of them. Another limitation is that the value of the citing books or articles was not assessed because this needs a separate qualitative study and extensive manual checking. It is therefore possible that some of the citations were from low value documents, especially in Google Scholar, such as student assignments. In addition, the method used to extract and filter citations from Google Books is likely to underestimate the citations received by highly cited books or books that for some reason generate many false matches (Kousha, Thelwall, & Rezaie, 2011). Finally, the coverage of Google Books and Google Scholar may change daily but seem unlikely to decrease and so the results should be taken as lower estimates.

The results show that the combination of Google Books and Google Scholar can provide good impact data for the books published by university publishers in Malaysia, in the sense that just under half of the tested books, including many recently published, had received at least one citation from one or both of these. Excluding recent books, then, it would be reasonable to expect a majority of Malaysian books to be cited at least once and hence it should be possible to differentiate books that have had some impact from books that have had none. The total number of citations was quite low overall, however, and skewed, so it would not be possible to make many finer-grained distinctions, except perhaps for identifying a small percentage of the most highly cited books. Hence the combination of Google Books and Google Scholar would be sufficient for purposes that primarily aimed to distinguish between work that has had some measurable impact and other work in a way that allowed about half of the publications to be regarded in and undifferentiated way as low impact work (e.g., placed in the lowest category, if a category-based scheme were to be used).

The finding that Malaysian books were less often cited than English books from Malaysia echoes previous findings of citation biases towards English documents (Basu, 2006; Seol & Park, 2008; Kellsey & Knievel, 2004). It is not clear whether the citation advantage of Malaysian English books is because they are higher quality or for other reasons, however, but this is additional evidence that academic publication in English should be encouraged, other factors being equal. A possible explanation for the apparent citation advantage of English articles is that authors knowing or believing that they can attract an

international audience would naturally not publish in a local language, and this would lead to more successful authors tending to publish in English. Previous studies have found a citation advantage for English publications in traditional citation databases (e.g., Van Leeuwen, Moed, Tijssen, et al., 2001) due to the *coverage* of those databases, which may be a factor also for Google Books and Google Scholar.

Conclusions

Based upon the above discussion, the use of Google Scholar and Google Books in conjunction is recommended for non-Western countries wishing to conduct impact assessment exercises for their locally-published social science or arts and humanities academic monographs or edited books. Further testing should be conducted for countries that are significantly less wealthy than Malaysia, which rarely publish in English, or which use a script that may be difficult to automatically scan, however. Although the coverage of Thomson Reuters BKCI and Scopus is presumably weak for non-Western countries, results from these could also be tested and added if improved coverage is needed. The results of such an assessment should be able to distinguish between about half of the books that have had some citation impact and the rest that have not. The results should also be able to identify a small percentage of highly cited books but would probably not be able to make many fine-grained distinctions between middle-ranking books since most cited books may have just a single citation.

The results can also be used to help identify the relative impact of different academic presses within a country. In the case of Malaysia, academic press impact did not follow university reputation, at least as indicated by the QS University Rankings, and so the citation scores books published by a press may be useful information for librarians choosing books and for research committees assessing books, especially if they are too newly published to have attracted citations.

For future work it would be useful to use the same methods to assess publications from poorer nations and from nations with scripts that are difficult to automatically scan.

Acknowledgement

References

Adler, N. & Harzing, A.W. (2009). When knowledge wins: transcending the sense and nonsense of academic rankings. *Academy of Management Learning and Education*, 8 (1), 72-95.

Aguillo, I, Bar-Ilan, J. Levene, M & Ortega, J.L. (2000). Comparing university rankings. Scientometrics 85 (1), 243-256.

Bar-Ilan, J. (2010). Citations to the "Introduction to informetrics" indexed by WOS, Scopus and Google Scholar. *Scientometrics*, 82(3), 495-506.

Basu, A. (2006). Using ISI's "highly cited researchers" to obtain a country level indicator of citation excellence. *Scientometrics,* 68 (3), 361-375.

Brown, L., Griffiths, R. & Rascoff, M. 2007. University publishing in a digital age. *The Journal of Electronic Publishing,* 10 (3), DOI: http://dx.doi.org/10.3998/3336451.0010.301.

Butler, L. & Visser, M.S. (2006) Extending citation analysis to non-source items. *Scientometrics*, 66 (2), 327-343.

Costas, R. & van Bochove, C. (2012). On the relationship between author collaboration and impact of scientific publications. Proceedings of 17th International Conference on Science and Technology Indicators, Montréal: Science-Metrix and OST, 447-448.

Cronin,B & La Barre, K. (2004). Mickey Mouse and Milton: Book publishing in the humanities. *Learned Publishing,*17 (2), 85-98.

Cronin, B., Snyder, H. & Atkins, H. (1997).Comparative citation rankings of authors in monographic and journal literature: a study of sociology. *Journal of Documentation* 53 (3), 263-273.

Cullars, J. (1992). Characteristics of the monographic literature of British and American literary studies.*College and Research Libraries*, 46 (1), 511–522.

Dassa, M., Kosmopoulos, C. & Pumain, D. (2010). JournalBase - A comparative international study of scientific journal databases in the social sciences and the humanities (SSH). *Cybergeo: European Journal of Geography,* Document 484, DOI : 10.4000/cybergeo.22862

García-Pérez, M.A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in psychology. *Journal of the American Society for Information Science and Technology*, 61(10), 2070–2085

Google (2013a). Inclusion guidelines for webmasters. Available at: http://www.google.com/intl/en/scholar/inclusion.html

Google (2013b). Google Scholar Citations: Available at: http://www.google.com/intl/en/scholar/citations.html#citations

Google (2013c). Where do these books come from?: Available at: https://support.google.com/books/answer/43726?hl=en-GB&ref_topic=9259

Google (2013d). What about books in other languages?: Available at: https://support.google.com/books/answer/43745?hl=en-GB&ref_topic=9259

Gump, S.E. (2006). Prestige and the university press. *Journal of Scholarly Publishing,* 37 (2), 69-85.

Hammarfelt, B. (2011). Citation analysis on the micro level: The example of Walter Benjamin's *Illuminations. Journal of the American Society for Information Science & Technology,* 65 (5), 819-830.

Harzing, A.W. (2010). *The Publish or Perish Book: Your guide to effective and responsible citation analysis*. Melbourne, Australia: Tarma Software Research Pty Ltd.

Holmes, R. (2006). The THES university rankings: are they really world class? *Asian Journal of University Education*, 2 (1), 1-14.

Jordy, M.L., McGarth, E.L. & Rutledge, J.B. (1999). Book reviews as a tool for assessing publisher reputation. *College & Research Libraries*, March, 132-142.

Kellsey, C. &Knievel, J.E. (2004). Global English in the humanities? A longitudinal citation study of foreign-language use by humanities scholars. *College & Research Libraries,* 65 (3), 194-204.

King, D. (2004). The scientific impact of nations, *Nature*, 430(6997), 311-316.

Kousha, K., Thelwall, M. (2009) Google Book Search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, 60 (8), 1537-1549.

Kousha, K., Thelwall, M. & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology,* 62(11), 2147–2164.

Levitt J.M. & Thelwall M. (2010). Does the higher citation of collaborative research differ from region to region? A case study of Economics. *Scientometrics* 85(1), 171–183.

Leydesdorff, L. & Wagner, C. (2008). Is the United States losing ground in science? A global perspective on the world science system. *Scientometrics,* 78(1), 23-36.

Lindholm-Romantschuk, Y. (1996). The role of monographs in scholarly communication: an empirical study of philosophy, sociology and economics. *Journal of Documentation,* 52 (4), 389-404.

Linmans, A J M. (2010). Why with bibliometrics the humanities does not need to be the weakest link - indicators for research evaluation based on citations, library holdings, and productivity measures. *Scientometrics*, 83 (2), 337-54.

Mingers, J. & Lipitakis, E. A. E. C. G. (2010). Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics,* 85 (2), 613–625

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics,* 66 (1), 81-100

Rauhvargers, A. (2011). Global university ranking and their impact. *EUA Reports on Ranking 2011*. Brussels, Belgium: European University Association.

Saenger, P. (1975). Colard Mansion and the evolution of the printed book. *The Library Quarterly*, 45 (4), 405-18

Seol, S.S. & Park, J.M. (2008). Knowledge sources of innovation studies in Korea: A citation analysis. *Scientometrics*, 75(1), 3-20.

Serenko, A., Bontis, N. & Moshonsky, M. (2012). Books as a knowledge translation mechanism: Citation analysis and author survey. *Journal of Knowledge Management*, 16 (3), 495-511.

Sieber, J. & Gradmann, S. (2011). How to best assess monographs? An attempt to assess the impact of monographs using library infrastructure and Web2.0 tools. WP6 Report. *European Educational Research Quality Indicators*. 23 p.

Shelton, R. D., Foland, P., & Gorelsky, R. (2009). Do new SCI journals have a different national bias? *Scientometrics*, 79 (2), 351-363.

Steele, C. (2008). Scholarly monograph publishing in the 21st Century: The future more than ever should be an open book. *The Journal of Electronic Publishing,* 11 (2). Available at: http://quod.lib.umich.edu/j/jep/3336451.0011.201?rgn=main;view=fulltext

Tausch, A. (2011). On the global impact of selected social-policy publishers in more than 100 countries. *Journal of Scholarly Publishing*, 42(4), 476-502.

Thompson, J. W. (2002). The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship, *Libri,* 52 (3), 121–136.

Torres-Salinas, D. & Moed, H. (2009). Library catalog analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics. *Journal of Informetrics,* 3 (1), 9-26

Torres-Salinas, D., Rodríguez-Sánchez, R., Robinson-Garcia, N., Fdez-Valdivia, J. & Garcia, J.A.(2013). Mapping citation patterns of book chapters in the Book Citation Index. *Journal of Informetrics*, 7 (2), 412-424.

Torres-Salinas, D., Robinson-Garcia, N., Jimenez-Contreras, E. & Lopez-Corza, E.D. (2012). Towards a book publishers citation reports: First approach using the "Book Citation Index".*Revista Espanola De Documentacion Cientifica*, 35(4), 615-624

Van Leeuwen, T.N., Moed, H.F. Tijssen, R.J.W., Visser, M.S., & Van Raan, A.F.J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics* 51(1), 335-346

Vaughan, L. & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes, Information Processing & Management, 40(4), 693-707.

Williams, P., Stevenson, I., Nicholas, D., Watkinson, A. & Rowlands, I. (2009). The role and future of the monograph in Arts & Humanities research, *Aslib Proceedings: New Information Perspectives*, 61 (1), 67-82.

White, H., Boell, S.K, Yu, H., Davis, M., Wilson, C.S. & Cole, F.T.H. (2009). Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences, *Journal of the American Society for Information Science and Technology*, 60 (6),1083-1096

Zhou, P. & Leydesdorff, L. (2006). The emergence of China as a leading nation in science Research Policy, 35(1), 83-104.

Zuccala, A. & van Leeuwen, T. (2011). Book reviews in humanities research evaluations, *Journal of the American Society for Information Science and Technology*, 62 (10), 1979-1991.