Comparing book citations in humanities journals to library holdings: Scholarly use versus 'perceived cultural benefit'.

Alesia Zuccala[1] and Raf Guns[2]

[1] *a.zuccala@uva.nl*
Institute for Logic, Language and Computation, University of Amsterdam,
P.O. Box 94242, Amsterdam, 1090 GE (The Netherlands)

[2] *raf.guns@ua.ac.be*
University of Antwerp, IBW, City Campus, Venusstraat 35, B-2000 Antwerpen (Belgium)

Abstract
In this paper we examine the statistical relationship between citation counts to books referenced in SCOPUS humanities journals and library holding counts ('libcitations') retrieved from WorldCat®. Our focus is on books (with ISBN numbers) published between 2001-2006, which received citations in *History* and *Literature & Literary Theory* journals during the period of 2007-2011. A Spearman's rank correlation coefficient was used, and our test resulted in significant correlations between the citations and 'libcitations'. We present and discuss the details of our dataset (extracted from a much larger, newly constructed database) and comment on why the 'perceived cultural benefit' of holding a book in a research library can lead to, but may not necessarily lead to use (i.e., a citation) of that book in new humanities research.

Conference Topic
Scientometrics Indicators: Relevance to Humanities (Topic 1), Old and New Data Sources for Scientometric Studies: Coverage, Accuracy and Reliability (Topic 2), and Bibliometrics in Library and Information Science (Topic 3).

Introduction
Books or monographs published in the humanities capture the research efforts of scholars concerned with human achievements. These texts are as much a part of our cultural heritage as they are part of scholarship (Garfield, 1979). In books we observe the story of a research discipline, that is, how it has evolved in different regions, over a specific time period, and within a particular "interpretive" community (Fish, 1980). Despite the fact that books are, for many humanities topics, principal modes of output, little is known about their scholarly *impact*. Bibliometricians have been reluctant to approach the subject of impact, because it is normally associated with high citation counts to and from articles published in scientific journals covered by the Web of Science (e.g., the Journal Impact Factor). Some journals published in the humanities are agreeable to impact factors (see Elsevier, 2010), but for the most part, these measures have been avoided in favour of general citation monitoring (Nederhof, 2006). Since the late 1970s, research has focused primarily on the *characteristics* of cited works in humanities texts or *classifying citations* to or from small monograph collections in disciplines where they are most prevalent (Budd, 1986; Cullars, 1985; 1989; 1998; Frost, 1979; Hammarfelt, 2012; Heinzkill, 1980; Hellqvist, 2010; Jones et al., 1972; Lindholm-Romantschuk & Warner, 1996; Nolan, 2010; Stern, 1983; Thomson, 2002).

For books in general, the absence of source metadata (i.e., internal identification codes) in the main commercial citation indices (i.e., Thomson Reuters' Web of Knowledge and Elsevier SCOPUS) has made it difficult to develop reliable indicators. Books have always been recorded in the Thomson Reuters' Web of Science (i.e., Science-SCI-E, Social Science-SSCI and Arts & Humanities-A&HCI) as 'non-sourced' cited materials, but some 'book chapters' and 'books' started to appear in all three indices as far back as 2005. Growth rates indicate

that their appearance has occurred irregularly (Leydesdorff & Felt, 2012). Recently, Leydesdorff & Felt (2012) found that the classification of books in the Web of Science is problematic: many have been misclassified as articles or reviews. Thomson Reuters' new Book Citation Index (BKCI) is expected to be a more accurate resource for bibliometric analyses (Adams & Testa, 2011). With the introduction to this index, we have been promised a 'complete picture' (Thomson Reuters, 2013). Only research based on this new Book Citation Index can tell us how useful it will be for evaluating citation-based impacts over the long term.

While the new indices are still in production, some researchers from the bibliometrics community have been considering alternative ways to study the impact of books. Kousha and Thelwall (2009) confirm that there are substantial numbers of citations to academic books from Google Books and Google Scholar to help evaluate book-oriented disciplines. Torres-Salinas and Moed (2009) as well as White et al. (2009) have focused on the potential of library catalogues for impact-based analyses, where an analogy may be created between journal-based citations and library holdings. Torres-Salinas and Moed (2009) studied the number of catalogue inclusions per book title in WorldCat®, while White et al. (2009) introduced the term 'libcitation' as "an indicator of perceived cultural benefit" (p. 1087). Linmans (2010) later suggested that researchers use a three-level approach for assessing books, focusing on citation counts, library holdings, and productivity.

The present study is motivated by the contributions of Torres-Salinas and Moed (2009) and White et al. (2009). Our objective is to further this earlier work using a special database that we have constructed to include books cited in journal articles covered by SCOPUS (*History* and *Literature & Literary Theory*) and corresponding library holding counts in both *Association of Research Libraries* (ARL) and non-ARL libraries. These were gleaned from WorldCat®. ARL is a non-profit membership organization of 125 research libraries in North America. Here, we explain how this database was developed for a much larger project (i.e., still a research in-progress) and present some preliminary analyses pertaining to the scholarly use of books (i.e., cited in journals) and their 'perceived cultural benefit' (catalogued in ARL and non-ARL libraries).

Overview of the datasets and database

Data were granted to us from Elsevier through the Elsevier Bibliometrics Research Program. In our application to this program we requested two separate datasets, each limited to citations recorded in journals classified as *History* or *Literature & Literary Theory* (Table1).

Table 1. Journals and journal citation data granted by Elsevier SCOPUS (April 2011).

Journal article publication years: (1996-2000 and 2007-2011)	
Journal Numbers and Classifications	
• History (n=604)	• ASJC 1202 (SCOPUS Classification Code)
• Literature & Literary Theory (n=529)	• ASJC 1208 (SCOPUS Classification Code)

Upon receiving the SCOPUS data, we examined the number of citations recorded in the 1023 journals (two time periods together) to determine the overall frequency to books, research articles (ar), conference proceedings (cp), review papers (re), notes (no) and other non-sourced materials. Cited materials that did not have an internal SCOPUS identification number, or did not meet the criteria that we established for identifying other materials - e.g., a

non-sourced journal/proceedings article - were classified as a 'book'. All 'other' documents will be re-examined and classified at a later date.

Table 2. Number of citations according to document type.

CITING - TO-	CITED	NUMBER
Research Article (ar)	**Books**	**1,647,520**
Research Article (ar)	Other documents	1,563,831
Review (re)	**Books**	**1,162,461**
Review (re)	Other documents	852,464
Research Article (ar)	Research Article (ar)	133,531
Review (re)	Research Article (ar)	41,855
Conference Proceeding (cp)	Books	41,800
Conference Proceeding (cp)	Other documents	34,854
Research Article (ar)	Review (re)	30,493
Notes (no)	Books	27,341

For all items classified as a 'book' we performed a set of queries in WorldCat®, using an API developer key granted to us by the Online Computer Library Center (OCLC). This key allowed us to match cited titles with the titles recorded in one or more of the international libraries covered by WorldCat®. For every matched title (confirming that it was a book), we retrieved an OCLC identification number, ISBN number, publisher name, publisher location, and a library holding count for both ARL and non-ARL libraries. With all the journal citing-to-cited data obtained from Elsevier, including the results of the API queries, we created a unique new SCOPUS-WorldCat® relational database.

Table 3, below, presents some descriptive statistics resulting from queries made to our new database. Here we show the total number of documents cited by articles or reviews published in *History* and *Literature & Literary Theory* journals for two citation windows: 1996-2001 and 2007-2011. Some of these cited documents have been categorized as follows: a) sourced in SCOPUS only, b) non-sourced in SCOPUS, but matched in WorldCat®, or c) sourced in SCOPUS and matched in WorldCat®. In this paper, we are concerned with a subset of books that were non-sourced in SCOPUS, but matched in WorldCat®.

Table 3. Cited documents as SCOPUS sourced or non-sourced items.

	All Cited Docs	Sourced in SCOPUS only	Not in SCOPUS, but Matched in WorldCat®	In SCOPUS & Matched in WorldCat®	Not in SCOPUS or WorldCat®	Cited Docs w. Missing Values (?)
HISTORY						
1996-2001	882,155	6,945	303,048	368	564,773	7,021
2007-2011	2,858,005	117,789	806,985	2,251	1,915,002	15,978
LITERATURE						
1996-2000	198,606	815	75,840	139	120,445	1,367
2007-2011	1,395,917	36,737	504,721	1,546	845,561	7,352

Data analyses and results

The aim of this study is to statistically examine the relationship between citation counts to books in SCOPUS journals and WorldCat® library holding counts ('libcitations') for both

ARL and non-ARL libraries worldwide. We expect to find a strong positive relationship, where 'libcitations' or library holdings support or lead to journal citations, but not vice versa. The idea behind this hypothesis is that humanities scholars borrow books from their university/academic library and 'use' these books by reading and citing them in research articles/review papers. Time is required for a book to be published, marketed to and purchased by a library before it is cited; hence we focus on a book publication window of six years (2001-2006), followed by a journal citation window of five years (2007-2011). With respect to library holdings, we assume that a book published in 2001 would not have been added to one of the libraries at least until this date or sometime after (up to and including Nov. 2012).
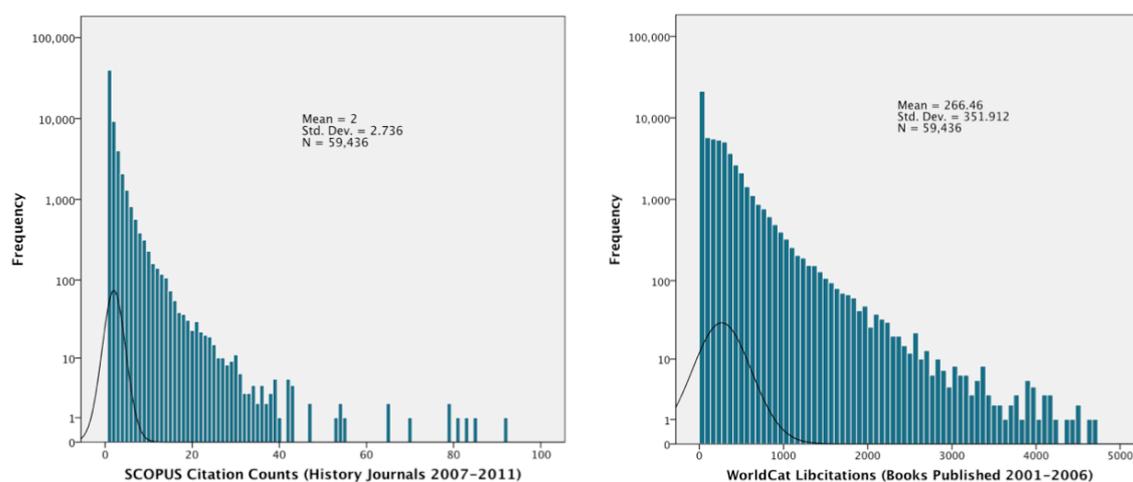

Figure 1. Citation and 'libcitation' frequency distributions for books published in 2001 to 2006 (cited in SCOPUS *History* journals, 2007-2011).

Table 4. Spearman's rank correlation coefficients for citations and libcitations.
(*History*, 2007-2011 and *Literature & Literary Theory*, 2007-2011).

PUBLISHED BOOKS (ISBN#)	NUMBER	CITATIONS SCOPUS JOURNALS	LIBCITATIONS ASSOCIATION OF RESEARCH LIBRARIES (ARL)	OTHER LIBRARIES (NON_ARL)	NON-PARAMETRIC CORRELATIONS SPEARMAN'S rho CITATIONS + ARL HOLDINGS	CITATIONS + NON_ARL HOLDINGS	CITATIONS + ALL LIBRARY HOLDINGS
HISTORY	59,436	Max. = 92 Min. = 1 Mean = 2.00	Max. = 212 Min. = 0 Mean = 46.36	Max. = 4,603 Min. = 0 Mean = 220.10	.288*	.267*	.275*
LITERATURE	41,853	Max. = 91 Min. = 1 Mean = 1.91	Max. = 215 Min. = 0 Mean = 48.35	Max. = 4,603 Min. = 0 Mean = 247.63	.281*	.244*	.254*
*correlation is significant at the 0.01 level (1-tailed).							

To select the best test for our hypothesis we first observed and compared the frequency distributions for all citation counts and library holding counts in the separate fields. Figure 1, above, presents the book citation and library holding distributions for *History* (*N*=59,436) only. Note that the data are skewed thus do not fit a normal curve. Similar non-normal

distributions were found for *Literature & Literary Theory* (*N*=41,853). With this data the most appropriate test to use is the Spearman's rank correlation coefficient. The procedure for performing a Spearman correlation is the same for a Pearson correlation; however, the Spearman rho is less sensitive to strong outliers. Table 4, above, presents some general statistics related to our datasets and indicates that we found significant, though not especially strong correlations between citations and 'libcitations' in *History* as well *Literature & Literary Theory*.

Discussion and Considerations for Further Research

First it is important to comment on the 'cleanliness' of the *History* and *Literature & Literary Theory* datasets. Since we were working with thousands of 'non-sourced' book references, it was necessary to examine repeat iterations of the reference strings to be sure that they were to the same book. For the most part, they were, but without a deep manual cleaning effort, we cannot say that the datasets were 'perfectly' clean. With respect to our correlation results, it is possible that if given access to cited book references from other books as well, the Spearman's rho might even be more significant. Also, there is much to be said about conducting this type of test, especially when correlation measures are not necessarily the best for understanding 'causes' and 'effects'. Many 'in between' variables can influence a correlation, some of which may be the browsing habits of humanities scholars, the concentrated nature of their work, and habit of citing books from their collegial network regardless of whether or not it is present in their institutional research library. Nevertheless, it is the goal of their institution to hold books that are 'perceived' to be beneficial to the culture of their research. From a bibliometric perspective, it is helpful to know if research-oriented libraries are doing for the humanities what they aim to do, which is to make quality books available for scholarly use.

A sizable portion of the books are present in many libraries but infrequently cited in the data set. The reverse (highly cited, but present in few libraries) is less common. When we examined the list of books that were proportionally cited least compared to their holding count, we discovered two main reasons for the divergence. First, reference works such as the *Oxford English Dictionary* have a very high 'perceived cultural benefit' but are not typically cited. The presence of this kind of book indicates that citations and 'libcitations' are not entirely interchangeable – they measure (partially) different dimensions. Second, the list contains many books that stem from other disciplines (e.g., Diagnostic and Statistical Manual of Mental Disorders). It seems likely that most books in this category would be highly cited if journals from their disciplines were part of the data set. More work may be done with larger datasets, for instance, expanding the data to include other humanities subjects, and/or making comparisons with cited books in the social sciences and sciences. There is also a strong opportunity here to further examine the role of book reviews, as 'gateway' documents, i.e., documents that encourage or discourage librarians to purchase books, and the motivation of scholars to read and cite books that were or were not selected based on reviews.

References

Adams, J., & Testa, J. (2011). Thomson Reuters Book Citation Index. In E. Noyons, P. Ngulube & J. Leta (Eds.), The 13th Conference of the International Society for Scientometrics and Informetrics (Vol. I, pp. 13-18). Durban, South Africa: ISSI, Leiden University and the University of Zululand.

Budd, J. (1986). Characteristics of written scholarship in American literature: A citation study. *Library and Information Science Research, 8,* 189–211.

Cullars, J. (1985). Characteristics of the monographic literature of British and American literary studies. *College & Research Libraries, 46,* 511-22.

Cullars, J. (1989). Citation Characteristics of French and German Literary Monographs. *Library Quarterly, 59,* 305-25.

Cullars, John. (1998). Citation characteristics of English language monographs in philosophy. *Library & Information Science Research, 20(1)*, 41–68.

Elsevier (2010). Latest impact factor figures from Elsevier's arts and humanities journals. Retrieved August 1, 2012 from http://about.elsevier.com/impactfactor/author-reports-93964/webpage/author-webpage-93964.html.

Fish, S. (1980). *Is there a text in this class? The authority of interpretive communities*. Harvard, MA: Harvard University Press.

Frost, C. O. (1979). The use of citations in literary research: A preliminary classification of citation functions. *Library Quarterly, 49,* 399-414.

Garfield, E. (1979) Is information retrieval in the arts and humanities inherently different from that in Science? The effect that ISI®'S Citation Index for the Arts and Humanities is expected to have on future scholarship. *The Library Quarterly, 50(1),* 40-57.

Hammarfelt, (2012). Following the footnotes: A bibliometric analysis of citation patterns in literary studies. Unpublished PhD Thesis, Uppsala University. Retrieved January 7, 2013 from

Heinzkill, R. (1980). Characteristics of references in selected scholarly English literary journals. *Library Quarterly, 50,* 352-365.

Hellqvist, B. (2010). Referencing in the humanities and its implications for citation analysis. *Journal of the American Society for Information Science and Technology, 61(2)*, 310–318.

Jones, C., Chapman, M., & Woods, P. Carr. (1972). Characteristics of the literature used by historians. *Journal of Librarianship, 4,* 137-56.

Kousha, K. & Thelwall, M. (2009) Google book citation for assessing invisible impact? *Journal of the American Society for Information Science and Technology, 60(8)*, 1537-1549.

Ledesdorff, L. & Felt, U. (2012). "Books" and "book chapters" in the book citation index (BKCI) and science citation index (SCI, SoSCI, A&HCI). *Proceedings of the American Society for Information Science and Technology, 49(1),* 1-7.

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

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

Nolan, D. S. (2010). Characteristics of la literatura: A reference study of Spanish and Latin American literature. *College & Research Libraries, 71(1),* 9-19.

Stern, M. (1983). Characteristics of the literature of literary scholarship. *College & Research Libraries, 44(4),* 199-209.

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

Thomson Reuters. (2013). Putting books back into the library: Completing the research picture. The Book Citation Index[SM]. Retrieved January 8, 2013 from http://wokinfo.com/products_tools/multidisciplinary/bookcitationindex/.

Torres-Salinas, D., & Moed, H.F. (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.

White, H., Boell, S.K, Yu, H., Davis, M., Wilson, C.S., & Cole, F.T.H. 2 (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.