Correlating Libcitations and Citations in the Humanities with WorldCat and Scopus Data

Alesia Zuccala[1] and Howard D. White[2]

[1] *spl465@iva.ku.dk*
Royal School of Library and Information Science, University of Copenhagen
Birketinget 6, DK-2300 Copenhagen S (Denmark)

[2] *whitehd@drexel.edu*
College of Computing and Informatics, Drexel University
32nd and Chestnut Streets, Philadelphia, PA, 19104 (USA)

Abstract

The term *libcitations* was introduced by White et al. (2009) as a name for counts of libraries that have acquired a given book. Somewhat like citations, these library holdings counts, which vary greatly, can be taken as indicators of the book's cultural impact. Torres-Salinas and Moed (2009) independently proposed the same measure under the name *catalog inclusions*. Both articles sought an altmetric for authors of books in, e.g., the humanities, since the major citation indexes, oriented toward scientific papers, have not served them well. Here, using very large samples, we explore the libcitation-citation relationship for the same books by correlating their holdings counts from OCLC's WorldCat with their citation counts from Elsevier's Scopus. For books cited in two broad fields of the humanities during 1996-2000 and 2007-2011, we obtain positive, weak, but highly significant correlations. These largely persist when books are divided by main Dewey class. The overall results are inconclusive, however, because the Scopus citation counts for the books tend to be very low. Further correlational research should probably use the much higher book citation counts from Google Scholar. Nevertheless, a qualitative analysis of widely held and widely cited books clarifies the libcitation measure and helps to justify it.

Conference Topic

Indicators

Introduction

Journal-oriented scientists have long had citation counts as an indicator of the impact of their articles, and journal-based citation indexes cater to them. But the same indexes cover citations to books less well, and book-oriented scholars in the humanities and softer social sciences feel themselves at a disadvantage, especially if citation measures are going to be used in performance evaluations and funding decisions (see Kousha, Thelwall, & Rezaie 2011 for a review). White et al. (2009) responded to this lack by proposing that one measure of a book's cultural impact could be the number of libraries that hold it. The idea behind this altmetric was that librarians who acquire a book are somewhat like scholars who cite it, in that both acts involve assessment and choice on behalf of communities of readers. To bring out the parallel, White et al. called the librarians' formal act of acquisition a *libcitation* (first syllable as in "library"). They wrote that the libcitation count (also known as a library holdings count) for a particular book "increases by 1 every time a different library reports acquiring that book in a national or an international union catalog. Readers are invited to think of union catalogs in a new way: as 'librarians' citation indexes'" (p. 1084). OCLC's WorldCat was mentioned as a prime example of a union catalog—that is, one that pools the cataloging records of OCLC member libraries and reports how many of them hold each cataloged item.

At the same time and wholly independently, Torres-Salinas and Moed (2009) made an identical proposal. Their name for libcitations (our term here) was *catalog inclusions*, and they, too, stressed the parallel between such inclusions and citations to journal articles (p. 11). They, too, named WorldCat as a potential source of library holdings data. Moreover, both

they and White et al. raised the possibility of empirically testing the relationship between libcitation counts and citation counts for the same set of books: are the two correlated?

The question is important because citation counts, when scrupulously used, have become a standard performance indicator in many disciplines, and, given the inadequacies of citation data for books, it would be very interesting if libcitations could serve a similar purpose. Torres-Salinas and Moed (2009, p. 24) saw correlation research of this sort in terms of validating the holdings-count idea:

> One way of doing this is to examine...the degree of correlation between the number
> of times book titles are cited in the serial literature on the one hand, and the number
> of library catalogs in which they are included on the other.

That is just what the present paper does for books (aka titles) in two broad fields in the humanities: *History* and *Literature & Literary Theory*. It draws on a special database of book citation data from Elsevier's Scopus and libcitation data for the same books from WorldCat, as described in Zuccala and Guns (2013), a research-in-progress paper. White et al. (2009, p. 1094) had anticipated what would be found:

> It is an open question whether libcitation counts for books and book chapters will
> correlate significantly with citation counts for the same works. Indeed, they may not.
> Our exploratory trials have shown some books to be high in both citation and
> libcitation counts. Occasionally, a book turns up that is well cited despite being held
> by relatively few libraries. More common are books that are meagerly cited, but
> relatively widely held. This overall mix produces low correlations.

These remarks were occasioned by spot-checking citation counts in the Web of Science. Using Scopus instead, Zuccala and Guns (2013) provided the first empirical answer to the open question: they found low but significant correlations.

The present paper continues this line of analysis (also described in Sieber and Gradmann, 2011). We do not hypothesize that libcitations *cause* citations (or the reverse)—merely that the two variables may positively co-vary.

Our database covers more than 100,000 books, and it now allows correlations to be obtained in the 10 main Dewey subject classes. As before, it has a total libcitation count for each book, but also disaggregates that total into counts for members of the Association of Research Libraries (ARL) and counts for non-members. The non-members include thousands of academic and public libraries whose collections are not primarily intended to support advanced research. In contrast, the 125 ARL institutions own very large subject collections that support graduate degree programs and specialized faculty research in many disciplines. (When multiple libraries in ARL institutions buy the same book, its count can go well beyond 125.) The books with the greatest cultural impact achieve libcitation counts in the thousands by appealing to ARL members and non-members alike. Plum Analytics, a commercial firm specializing in altmetrics, now includes a book's holding count in WorldCat as one of its indicators of "usage."

The results of our analyses, while interesting and suggestive, return us to a common criticism of both the Web of Science and Scopus: within the time frame of our study, they pick up too few citations to books to correlate those citations with libcitations on a firm basis. Both WoS and Scopus have recently expanded their efforts to capture citations to books, but it is too early to assess the full effect of these new data on bibliometrics. Kousha, Thelwall and Rezaie (2011) demonstrate that Google Books and Google Scholar give considerably higher citation counts for books than Scopus does. Our findings point to the same conclusion.

Overview of the database

Here we re-present several details about our database from Zuccala and Guns (2013) and add some new ones. The Scopus database from Elsevier supplied our citation data, which was

granted through the Elsevier Bibliometrics Research Program. Having requested separate datasets in *History* and *Literature & Literary Theory*, we further limited them to citations that appeared in journal articles during two periods, 1996-2000 and 2007-2011. We examined the Scopus data to determine the overall frequency with which various types of publications were cited: books, research articles, conference proceedings, review papers, notes, and other materials. Cited materials that were "non-sourced"—that is, that did not have a Scopus identification number linking them to a source journal—were classified as books, the unit of analysis in which we were interested.

Table 1 shows the number of journals in each field (as classified by Scopus) from which we drew citing articles. The lower part of Table 1 gives the numbers (N's) of books cited in the journal articles in each field and period. It will be seen that, in both fields, the N's of books cited in the earlier period are much smaller than those in the later, because Scopus covered fewer humanities articles in the 1990s.

Table 1. Journals and journal citation data in Scopus (April 2011).

Journal counts and classification codes	
History (N=494 source journals)	ASJC 1202 (Scopus Classification Code)
Literature & Literary Theory (N=419 source journals)	ASJC 1208 (Scopus Classification Code)
Both History and Literature (N=110 source journals)	Both ASJC 1202 and ASJC 1208
Counts of books cited during 1996-2000	*Counts of books cited during 2007-2011*
History (N=20,996)	History (N=50,466)
Literature & Literary Theory (N=7,541)	Literature & Literary Theory (N=35,929)

We searched the apparent books in WorldCat, using an API developer key granted to us by the Online Computer Library Center (OCLC). The key allowed us to match titles cited *at least once* in Scopus with titles held by *at least one ARL and one non-ARL library* covered by WorldCat. (These libraries, while mostly North American, include participants worldwide.) For every matched title (confirming that it was a book), we retrieved the OCLC accession number, ISBN number, publisher's name, publisher's location, and library count data. These were added to the book's citation data from Scopus to create a unique Scopus-WorldCat relational database.

Once a book has been published, it takes time for it to be acquired and cataloged by a library. A book published in a given year could have been acquired by a library no earlier than that year, but might have been acquired up to and including November 2012. Our holdings counts were current as of that cut-off date.

To improve publication-date accuracy, we analyzed only books published in the six years immediately preceding our two five-year citation windows. Thus, the books cited in 1996-2000 were limited (by filtering their Scopus records) to those published during 1990-1995. The books cited in 2007-2011 were likewise limited to those published during 2001-2006.

Converted to the four files at the bottom of Table 1, our book data come to 114,932 cases in all, 81 percent of which are unique titles. The remaining 19 percent are titles that appear more than once. Some were cited in both our earlier and later periods. Others were cited in both the History and the Literature journals, or in the journals that Scopus has assigned to both fields jointly, as shown in Table 1. We did not attempt to re-assign these latter titles to a single field, but allowed them to enter into the counts for both fields. There seems no easy way to avoid double counting, because that is the way in which Scopus has structured the data. Even so, a trial analysis with duplicates removed does not greatly affect the correlations.

Data analyses and results

Our data analyses were conducted with SPSS, the Statistical Package for the Social Sciences. Table 2 gives summary statistics for the titles in History and Literature. Means and standard deviations have been rounded to whole numbers. As noted in Zuccala and Guns (2013, p. 357), both citations and libcitations exhibit the highly skewed distributions typical of bibliometrics. However, the subsets of ARL libcitations for both History and Literature have bimodal distributions, with peaks at 1-4 and 100-104 holding libraries, and a low point at 45-54 libraries. In other words, the ARL libraries tend to acquire large numbers of rarely held titles, large numbers of widely held titles, and markedly fewer titles with holdings counts in between. This saddle-shaped distribution may reflect the opposing needs of specialized researchers: on their behalf, ARL libraries acquire many books held by few other members, but also many books that almost every member *must* have. The titles with the maximum counts in Table 2 (e.g., 92 citations; 4,725 libcitations) will be named in Tables 6 through 9.

Table 2. Summary statistics for two fields in combined time periods.

History combined periods N=71462					
	Minimum	Maximum	Mean	Std. Dev.	Median
Citations	1	92	2	3	1
ARL libcitations	1	212	59	40	63
Non-ARL libcitations	1	4603	278	351	178
Total libcitations	2	4725	338	372	250
Literature combined periods N=43470					
	Minimum	Maximum	Mean	Std. Dev.	Median
Citations	1	91	2	3	1
ARL libcitations	1	215	62	38	67
Non-ARL libcitations	1	4603	305	395	189
Total libcitations	2	4725	367	412	267

In Table 3, *citation* counts for every book are correlated with total *libcitation* counts for every book in major subsets of the database. Citation counts are also separately correlated with the libcitation counts for ARL members and non-members. (Only the libcitation variables are labeled, but the unlabeled citation variable is present in all the cells.) These are Spearman rho correlations, calculated with ranks of the count values rather than the counts themselves. Unlike Pearson r's, rho's do not require the assumption of normally distributed populations and so accommodate bibliometric skew (Zuccala & Guns, 2013: 357).

Table 3. Total, ARL, and non-ARL libcitations to books correlated with citations to the same books in two fields, two periods, and combined periods.

History 1996-2000			History 2007-2011			History combined		
Total	ARL	Non-ARL	Total	ARL	Non-ARL	Total	ARL	Non-ARL
0.26	0.29	0.25	0.25	0.28	0.24	0.24	0.26	0.23
N=20996			N=50466			N=71462		
Literature 1996-2000			Literature 2007-2011			Literature combined		
Total	ARL	Non-ARL	Total	ARL	Non-ARL	Total	ARL	Non-ARL
0.23	0.28	0.22	0.18	0.24	0.17	0.20	0.24	0.19
N=7541			N=35929			N=43470		

The rho's are all positive and weak, with values much like those in Zuccala and Guns (2013, p. 357). Because of the large numbers of books involved, all are significant at p < .001 by

one-tailed test. The hypothesis of no relationship can thus be safely rejected: citations and libcitations do capture a certain amount of scholarly impact in common. A sign of this in Table 3 is that citations, which are essentially a researchers' practice, always correlate a bit more highly with libcitations from research libraries—that is, ARL members. However, none of the rho's are strong enough to indicate that libcitations can substitute for citations as a measure. Libraries, especially ARL members, do buy many books that turn out to be well cited, but they buy even more books that are not highly cited in the journals covered by Scopus. This raises questions about the citation-libcitation relationship that we will return to later with specific examples.

Table 4 may clarify the situation in our two subject fields. The total libcitation counts for books have been divided at their medians. Citation counts for the same books have been collapsed into three groups, as shown in the column labels. In both History and Literature, the two variables are directly related: as citation counts rise, the percentage of books with above-median libcitation counts also rises sharply. For example, in History, only 43% of books cited once have libcitation counts in the top half, whereas for books cited two to four times the comparable figure is 59%, and for books with five or more citations, 79%. The percentages in the Literature table are almost identical.

Table 4. Libcitations and citations cross-tabulated in two fields for combined periods.

History combined periods				
	Citations			
Libcitations	1	2-4	5 or more	
GT Median	43%	59%	79%	50%
LE Median	57%	41%	21%	50%
	100%	100%	100%	100%
N =	46578	19165	5719	71462
Literature combined periods				
	Citations			
Libcitations	1	2-4	5 or more	
GT Median	44%	59%	78%	50%
LE Median	56%	41%	22%	50%
	100%	100%	100%	100%
N =	29876	10668	2926	43470

However, this effect must be viewed in light of the extreme skew of the citation counts seen in the column marginals. Roughly two-thirds of all books in our samples have only one citation each, and roughly another quarter have only two to four citations. The fraction of titles with five or more citations is relatively small. Thus, the Spearman rho's for these grouped variables, though highly significant ($p < .001$), are even lower than when the variables are ungrouped in Table 3—only 0.22 for History and 0.19 for Literature.

We turn to a finer breakdown of the data. As mentioned in Zuccala and Guns (2013, p. 358), historians who publish in History journals do not exclusively cite works of history, nor do literary scholars who publish in Literature journals exclusively cite works of literature or literary theory. Instead, both groups cite books across the full range of subjects covered by the Dewey Decimal Classification. We were able to get the Dewey class numbers for most of our book titles from WorldCat. (Some books do not receive Dewey classifications.) In Table 5 we subdivide the books cited in History and Literature journals in our two time periods by their main Dewey classes.

Class 000 in Dewey is formally "Computer science, information, general works." This class is traditionally used for general reference books and books in trans-disciplinary fields such as librarianship, journalism, publishing, and reading. Historians and literary scholars mainly cite

books in areas like these, rather than in computer science. Hence, we have shortened the long label here to "General works."

The Table 5 cells contain 120 replications of our correlational study in subsets of the data. We are again correlating each book's total citations with its total libcitations, as well as the libcitation counts from ARL members and ARL non-members. In making comparisons, be aware that non-ARL libcitations make up by far the larger share of total libcitations. The two categories thus tend to produce correlations that are similar or identical, and so the non-ARL results will not be separately discussed here.

Table 5. Libcitations correlated with citations to books by field, period, and main Dewey classes.

	History 1996-2000				History 2007-2011			
Main Dewey Classes	Libcites	ARL	Non-ARL	N =	Libcites	ARL	Non-ARL	N =
000 General works	0.20	0.21	0.20	350	0.23	0.28	0.22	794
100 Philosophy and psychology	0.20	0.21	0.19	1055	0.18	0.20	0.17	2041
200 Religion	0.27	0.27	0.26	1766	0.27	0.29	0.25	4186
300 Social sciences	0.26	0.28	0.26	8067	0.23	0.25	0.21	16585
400 Language	0.11	0.11	0.12	247	0.17	0.16	0.17	672
500 Science	0.20	0.27	0.19	914	0.13	0.23	0.11	1543
600 Technology	0.25	0.35	0.23	824	0.12	0.24	0.09	1990
700 Arts and recreation	0.21	0.24	0.20	1056	0.19	0.26	0.18	3788
800 Literature	0.17	0.26	0.15	1620	0.20	0.26	0.19	4725
900 History and geography	0.28	0.31	0.27	4388	0.27	0.29	0.25	10439
	Literature 1996-2000				Literature 2007-2011			
Main Dewey Classes	Libcites	ARL	Non-ARL	N =	Libcites	ARL	Non-ARL	N =
000 General works	0.09	0.08	0.09	155	0.17	0.36	0.14	548
100 Philosophy and psychology	0.19	0.22	0.18	585	0.23	0.27	0.22	1919
200 Religion	0.13	0.19	0.12	398	0.25	0.29	0.23	2221
300 Social sciences	0.14	0.16	0.14	1344	0.19	0.22	0.18	6322
400 Language	0.22	0.24	0.21	505	0.22	0.24	0.20	1218
500 Science	0.04	0.09	0.04	115	0.06	0.12	0.06	516
600 Technology	0.13	0.28	0.11	130	0.09	0.24	0.07	703
700 Arts and recreation	0.18	0.21	0.17	591	0.22	0.26	0.20	3268
800 Literature	0.23	0.31	0.21	2616	0.26	0.31	0.25	11171
900 History and geography	0.14	0.25	0.12	742	0.21	0.26	0.20	3963

Even with Table 5's extensive partitioning, the N's underlying the correlations are large enough that most of the rho's remain highly significant ($p < .001$ by one-tail test). Of the correlations between citations and total libcitations, 21 out of 40 remain at or above 0.20. Large N's can cause correlations that are statistically but not substantively significant (Babbie 2015, p. 469). Nevertheless, certain patterns do lend substance to the overall analysis:

- Some 33 of the 40 ARL correlations remain in the 0.20s or higher.
- Some 37 of the 40 ARL correlations are higher than those for the non-ARL libraries in their row. This reinforces the supposed connection between citations and libcitations in research environments.
- As examples of subject accord, the ARL correlation for books classed in *900 History and geography* is second-highest (0.31) in History 1996-2000, and tied-highest (0.29) in History 2007-2011.
- As further examples of subject accord, the ARL correlation for books classed in *800*

Literature is highest (0.31) in Literature 1996-2000, and second-highest (0.31) in Literature 2007-2011.

- In both our History periods, the lowest correlations occur for books classed in *400 Language*. The N's for books in this class, which is historically Dewey's smallest, are likewise small. While historians make use of research from all fields, it is unsurprising that books on language are not their chief resource.

- In both our Literature periods, the lowest correlations occur for books classed in *500 Science,* and the N's for books in this class are small as well. One would not expect literary scholars to cite numerous science books. However, one might expect them to cite more books in *400 Language* than historians, and that is what the data show.

- Table 5 in fact shows wide variation in the number of books that Scopus authors have cited in each class. In both History periods, books classed in *300 Social Sciences* are most numerous. This makes sense because of the close interplay between historical and social scientific topics. Books classed in *900 History and geography* are the second-most numerous, and books in *800 Literature* are third. In both Literature periods, the same three classes dominate but in another order: *800 Literature* first, as seems fitting, then *300 Social Sciences* and *900 History and geography*. For our two broad fields in the humanities, these are reassuringly reasonable outcomes.

Since libcitations are a new altmetric, we think it informative to display the titles that have top-ranked libcitation counts in particular contexts (as do both Torres-Salinas and Moed, 2009 and White et al., 2009). This allows a qualitative as well as a quantitative analysis. White (2005) proposed the label *bibliograms* for bibliometric distributions in which not only the ranked counts but also the terms associated with them are analyzed as communications. "Bibliograms," he wrote (p. 443), "consist of (1) at least one seed term that sets a context, (2) terms that co-occur with the seed across some set of records, and (3) counts of how frequently terms co-occur with the seed by which they can be ordered high to low." Here, we use main Dewey class names as seed terms. We then rank the books that co-occur with them (as OCLC accession numbers) by their libcitation or citation counts. Lastly, the OCLC numbers are used to retrieve full bibliographic data from WorldCat so that we can comment on the authors, titles, and nature of the top-ranked books.

Table 6 comprises extracts from 40 bibliograms. We display, for our two fields and two time periods, the titles with the highest *total* libcitation counts in each of the 10 main Dewey classes. Many of these books have subtitles, but they have been omitted in favor of authors' surnames (or those of first authors in collaborations). We also display their ARL libcitation counts and their citation counts in Scopus.

The books in Table 6 do not resemble typical scientific articles. They are the sort of titles that present readers, like everyone else, may have purchased for reasons having nothing to do with bibliometrics. They exemplify the broad cultural impact of the humanities—for example, standard reference works on language, music, religion; biographies of famous men (Peter Gay's *Freud*, David McCullough's *Truman* and *John Adams*); novels (Toni Morrison's *Paradise*, Dan Brown's *The Da Vinci Code*); popularizations of science (Dava Sobel's *Longitude*, Malcolm Gladwell's *Blink*, Carl Sagan's *Cosmos*); best-selling social critiques (Susan Faludi's *Backlash*, Robert Hughes's *Culture of Complaint*); advice for business executives (James Collins's *Good to Great*, Thomas Peters and Robert Waterman's *In Search of Excellence*). While some exemplify high scholarship, others are not scholarly at all (Ernest Hemingway's *A Moveable Feast*); some are even children's books (David Wiesner's *Flotsam*, Peter Spier's *Noah's Ark*, both Caldecott Medal winners). They come to the fore here because they were bought by thousands of libraries, and they had citation counts of at least one in Scopus. Persons at research universities who specialize in manifestations of popular culture are legion.

Table 6. Books with highest libcitation counts by field, period, and main Dewey class.

History 1996-2000					
Cites	**ARL**	**Libcites**	**Dewey class**	**Title**	**Author**
1	160	2592	General works	The Oxford dictionary of modern quotations	Augarde
1	143	2936	Philosophy and psychology	Freud	Gay
1	101	2789	Religion	Crossing the threshold of hope	John Paul II
1	124	4233	Social sciences	My American journey	Powell
1	105	3433	Language	The story of English	McCrum
2	108	2572	Science	Longitude	Sobel
1	112	3204	Technology	Healing and the mind	Moyers
1	130	2133	Arts and recreation	Culture of complaint	Hughes
1	122	4132	Literature	Paradise	Morrison
4	137	4724	History and geography	Truman	McCullough

History 2007-2011					
Cites	**ARL**	**Libcites**	**Dewey class**	**Title**	**Author**
1	160	2592	General works	The Oxford dictionary of modern quotations	Augarde
2	145	4059	Philosophy and psychology	Blink	Gladwell
1	93	2931	Religion	Under the banner of heaven	Krakauer
4	152	3967	Social sciences	Freakonomics	Levitt
5	182	2760	Language	The Oxford English dictionary	Simpson
4	104	3284	Science	A short history of nearly everything	Bryson
2	148	4496	Technology	In search of excellence	Peters
4	123	2596	Arts and recreation	New Grove dictionary of music	Grove
6	122	4725	Literature	The Da Vinci code	Brown
5	140	4655	History and geography	John Adams	McCullough

Literature 1996-2000					
Cites	**ARL**	**Libcites**	**Dewey class**	**Title**	**Author**
2	155	2076	General works	Double fold	Baker
3	145	4059	Philosophy and psychology	Blink	Gladwell
1	87	3511	Religion	Noah's ark	Spier
3	152	3967	Social sciences	Freakonomics	Levitt
1	105	3433	Language	The story of English	McCrum
1	125	3884	Science	Cosmos	Sagan
1	141	4195	Technology	Good to great	Collins
1	86	4133	Arts and recreation	Flotsam	Wiesner
13	122	4725	Literature	The Da Vinci code	Brown
1	140	4655	History and geography	John Adams	McCullough

Literature 2007-2011					
Cites	**ARL**	**Libcites**	**Dewey class**	**Title**	**Author**
1	115	3342	General works	The road ahead	Gates
1	75	2455	Philosophy and psychology	Care of the soul	Moore
1	128	3083	Religion	The Oxford companion to the Bible	Metzger
2	154	3169	Social sciences	Backlash	Faludi
2	148	3119	Language	The Oxford companion to the English language	McArthur
1	112	2068	Science	Black holes and time warps	Thorne
1	93	4314	Technology	Men are from Mars, women are from Venus	Gray
1	130	2133	Arts and recreation	Culture of complaint	Hughes
1	125	3455	Literature	A moveable feast	Hemingway
1	121	3600	History and geography	The fifties	Halberstam

Thus, even the most pop-cultural books in Table 6 are widely held by ARL members. It is a misconception that these libraries acquire only works of rarified scientific or scholarly status. In fact, they buy innumerable works that would also be found in public and school libraries. The best example is the single most widely held item in our database—*The Da Vinci Code*, owned by 122 (of 125) ARL members. Whatever one may think of this novel, it had a huge impact for several years, and scholars in the humanities will want copies on hand, if only to attack Dan Brown's transgressions. Nevertheless, the citation counts for these books in Table 6's leftmost column are very low. Brown's novel has the most, and these may include book reviews.

By contrast, Table 7 displays the titles that are *most* highly cited in our categories. As implied earlier, relatively high citation counts tend to signal a research orientation, and these 40 books, which have the top counts in their respective Dewey classes, are almost all distinctly more academic than those in Table 6. Their *total* libcitation counts tend to be lower than those in Table 6, suggesting more specialized readerships. (The exception is *The Guardian*, a Nicholas Sparks novel.) A fair number of them address themes prominent in the humanities (race, class, gender, imperialism), and their authors include names famous to postmodern scholars, if not to the general public (e.g., Edward Said, Gilles Deleuze, Judith Butler, Donna Haraway, Gayatri Spivak, and, with two books, Giorgio Agamben).

Three-fourths of these books are held by a hundred or more ARL libraries. Of those that are not, some may reflect genuinely narrower acquisition by ARL members. Others (if not errors) may reflect delayed or incomplete reporting of an acquired book that makes its libcitation count deceptively small. That may have happened, for instance, with Spivak's *Death of a Discipline*, whose ARL count in Table 7 is only 22, but whose count as an e-book in WorldCat is 1,246 at this writing.

In any event, ARL libcitation counts range unbrokenly over values from 1 to 215. Given this variation, why are the correlations of ARL counts with citations not higher? We have already noted that they tend to be higher than correlations of *total* libcitations with citations, but only slightly. In both cases the problem is the same: the great majority of books in our database have only one citation (or at most a few). Thus, a key variable in our study has little variability. As one illustration, Table 8 lists the five books with the highest ARL libcitation counts in our two fields (time periods combined, and omitting the *Oxford English Dictionary*, already shown). These books are best-sellers not only among ARL members but in libraries of all kinds. Yet their citation counts in Scopus are minuscule and much the same, just as they were for the books in Table 6. To anyone familiar with these titles, it is incredible that Table 8 reflects their full citation records. Rather, their true counts are not being captured.

Not too long ago, this assumption could only have been checked with data from the Web of Science, but now we can spot-check citations to books in Google Scholar. When that is done, the results are very different from what Scopus shows, whether the Scopus figures are as low as one or as high as 92. Table 9 suggests the nature of the problem. The counts there reflect our judgment calls, such as to include only those for the 2000 edition of *DSM-IV-TR* or the 2007 edition of *The Elements of Style*. Google Scholar itself does not break down by edition the many citations to the feminist classic *In a Different Voice*. Nor does it allow us to extract citations to books in our two periods of study. Nevertheless, the Google Scholar counts indicate where further correlational research should be directed (see also Prins et al., 2014).

Table 7. Books with highest citation counts by field, period, and main Dewey class.

History 1996-2000					
Cites	**ARL**	**Libcites**	**Dewey class**	**Title**	**Author**
14	117	573	General works	The letters of the Republic	Warner
30	115	798	Philosophy and psychology	The production of space	Lefebvre
19	111	689	Religion	Ritual theory, ritual practice	Bell
75	129	1195	Social sciences	Imagined communities	Anderson
11	76	509	Language	Biblical Hebrew syntax	Waltke
29	107	450	Science	Bayes or bust?	Earman
25	84	364	Technology	Curing their ills	Vaughan
13	108	650	Arts and recreation	Orientalism	MacKenzie
56	119	1381	Literature	Culture and imperialism	Said
71	119	1406	History and geography	Britons	Colley
History 2007-2011					
Cites	**ARL**	**Libcites**	**Dewey class**	**Title**	**Author**
24	114	546	General works	"The tyranny of printers"	Pasley
39	26	413	Philosophy and psychology	The navigation of feeling	Reddy
37	109	478	Religion	Formations of the secular	Asad
92	114	602	Social sciences	Carnal knowledge and imperial power	Stoler
22	12	481	Language	Bilingualism and the Latin language	Adams
31	115	556	Science	The body of the artisan	Smith
32	100	342	Technology	Contagious divides	Shah
17	92	412	Arts and recreation	The reformation of the image	Koerner
26	32	2802	Literature	The guardian	Sparks
83	116	813	History and geography	The birth of the modern world, 1780-1914	Bayly
Literature 1996-2000					
Cites	**ARL**	**Libcites**	**Dewey class**	**Title**	**Author**
71	110	415	General works	The reading nation in the Romantic period	St. Clair
79	102	391	Philosophy and psychology	The open	Agamben
36	87	404	Religion	Saint Paul	Badiou
91	117	545	Social sciences	State of exception	Agamben
37	101	377	Language	The translation zone	Apter
12	95	294	Science	The spacious word	Padron
37	71	259	Technology	The companion species manifesto	Haraway
27	104	348	Arts and recreation	In the break	Moten
85	22	559	Literature	Death of a discipline	Spivak
87	106	462	History and geography	Writing history, writing trauma	LaCapra
Literature 2007-2011					
Cites	**ARL**	**Libcites**	**Dewey class**	**Title**	**Author**
6	117	573	General works	The letters of the Republic	Warner
17	108	632	Philosophy and psychology	Difference and repetition	Deleuze
6	114	771	Religion	Fragmentation and redemption	Bynum
41	131	1049	Social sciences	Gender trouble	Butler
19	84	301	Language	Discourse and social change	Fairclough
9	117	1034	Science	The origins of order	Kauffman
5	112	475	Technology	The commodity culture of Victorian England	Richards
11	122	983	Arts and recreation	Gone primitive	Torgovnick
38	120	843	Literature	The location of culture	Bhabha
23	125	891	History and geography	Imperial eyes	Pratt

Table 8. Books with the top five ARL libcitation counts in two fields.

Cites	ARL	Libcites	Title	Author
History combined				
Cites	**ARL**	**Libcites**	**Title**	**Author**
2	212	4101	Diagnostic and statistical manual of mental disorders: DSM-IV-TR	
2	194	2478	In a different voice	Gilligan
3	180	1282	The alchemy of race and rights	Williams
2	176	1348	On the law of nations	Moynihan
1	176	1136	Theoretical perspectives on sexual difference	Rhode
Literature combined				
Cites	**ARL**	**Libcites**	**Title**	**Author**
1	215	3792	Publication manual of the American Psychological Association	
1	204	3436	The elements of style	Strunk, White
1	203	2046	A theory of justice	Rawls
3	178	1466	There's no such thing as free speech, and it's a good thing, too	Fish
1	175	995	Sex and reason	Posner

Table 9. Same data, but with citations in Scopus replaced by citations in Google Scholar.

Cites	ARL	Libcites	Title	Author
History combined				
Cites	**ARL**	**Libcites**	**Title**	**Author**
5364	212	4101	Diagnostic and statistical manual of mental disorders: DSM-IV-TR	
30044	194	2478	In a different voice	Gilligan
2431	180	1282	The alchemy of race and rights	Williams
146	176	1348	On the law of nations	Moynihan
102	176	1136	Theoretical perspectives on sexual difference	Rhode
Literature combined				
Cites	**ARL**	**Libcites**	**Title**	**Author**
1393	215	3792	Publication manual of the American Psychological Association	
2988	204	3436	The elements of style	Strunk, White
782	203	2046	A theory of justice	Rawls
616	178	1466	There's no such thing as free speech, and it's a good thing, too	Fish
1546	175	995	Sex and reason	Posner

Discussion

The correlations in this paper suggest that libcitations and citations are not entirely different measures of impact. However, we are left wanting citation counts for books that do not have so many low, tied values. It is possible that better data would again produce low or even negligible correlations. It is also possible that the correlations would be much higher than those seen here. The libcitation measure draws on a varied mix of assessments, and they are not necessarily the same as those that go into scholars' acts of citation. But, as our data make plain, they indicate major intellectual achievements no less forcefully than citations. In fact, one can argue that many of the humanities titles in Table 6 are *truly* major achievements, in that they have reached large publics beyond academe.

What, then, do libcitations measure? Briefly, they estimate the potential readerships, or users, of a given book. Citations, in contrast, measure actual uses to which the book has been put within research-oriented communities. It is therefore not surprising that citations and libcitations are associated, especially if the latter come from libraries that serve researchers,

such as those in ARL. But libcitations also measure broad cultural impacts that citations may miss, because libcitations rest on chains of judgments within the world of publishing, and this world, which subsumes the scholarly one, extends into every part of life. The chains include authors, agents, past editors who have built publishers' reputations, present-day editors of various kinds, referee-readers, marketers, and wholesalers. Librarians are only the last link.

This speaks to the common objection that librarians do not evaluate individual titles, but put their acquisitions on automatic pilot through approval plans and the like; how, then, can libcitations reflect genuine worth? On the contrary, librarians are highly attuned to potential demand in their communities, and it is they who approve the approval plans and buy into the pre-formed collections. It is quite true that such moves favor some publishers over others, but that is because librarians trust the chains of judgment those publishers represent. And so do their communities, who routinely expect librarians to have acquired certain books they learn about and are displeased if they have not.

Libcitations are sales figures—a market measure. They reflect virtual unanimity on the worth of some titles, but they vary enormously. In our database, although the counts run to the high values seen in our tables, many titles are held by only one ARL and one non-ARL library, just as many papers have only a citation or two. Research on libcitation-citation correlations should continue, but even if they remain low, that does not invalidate the libcitation measure. It is better thought of as a free-standing gauge of authors' cultural impact. Having published a book, what author would not prefer a thousand libraries to hold it rather than 10?

Acknowledgments

References

Babbie, E. R. (2013). *The practice of social research*. 14th edition. Boston, MA: Cengage Learning.

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.

Prins, A., Costas, R., van Leeuwen, T., & Wouters, P. (2014). Using Google Scholar in research evaluation of social science programs, with a comparison with Web of Science data. *Proceedings of the Science and Technology Indicators Conference 2014 Leiden*, 434-443.

Sieber, J., & Gradmann, S. (2011). How to best assess monographs? An attempt to assess the impact of monographs using library infrastructure and Web 2.0 tools. European Educational Research Quality Indicators. Retrieved December 17, 2014 from

http://edoc.hu-berlin.de/docviews/abstract.php?id=38002

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. D. (2005). On extending informetrics: An opinion paper. Proceedings of the 10th International Society for Scientometrics and Informetrics Conference, 2, 442-449.

White, H. D., 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.

Zuccala, A., & Guns, R. (2013). Comparing book citations in humanities journals to library holdings: Scholarly use versus 'perceived cultural benefit' (RIP). *Proceedings of the 14th International Society for Scientometrics and Informetrics Conference, 1*, 353-360.