Notes by Megan Fitzgibbons
October 26, 2010

Appraisal tool

Evaluation Tool for Bibliometric Studies
By Carol Perryman MSLIS, TRLN Doctoral Fellow, School of Information & Library Science, University of North Carolina at Chapel Hill, March 5, 2009

About the instrument: The initial structure is derived from an evaluation instrument for quantitative research1 and amended to incorporate criteria identified by Glanzel (1996)2. iThis document is intended for quite different purposes than those presumed by Glanzel, whose concern it is to support standardization in methodology and terminology between bibliometric researchers. This tool is meant to support critical evaluation of bibliometric research for non-experts in this subspecialty area, with the purpose of practical application of the results within the framework of Evidence-based Library and Information Practice. See http://www.box.net/shared/7d7ga8r9z6

1. Basic information
Citation of work being appraised Antelman, Kristin. (2004). "Do open-access articles have a greater research impact?" *College & Research Libraries, 65*(5), 372-382.
Stated objective(s) of research This study's hypothesis is that scholarly articles from disciplines with varying rates of open-access adoption have a greater research impact if the articles are freely available online than if they are not.
Key findings The data show a significant difference in the mean citation rates of open-access articles and those that are not freely available online in all four disciplines. The relative increase in citations for open-access articles ranged from a low of 45 percent in philosophy to 51 percent in electrical and electronic engineering, 86 percent in political science, and 91percent in mathematics.
Does the study use mixed methods (more than one kind of method)? If so, what other method(s) are used? You may need to use additional evaluative tools. Mainly: collecting citation rates from ISI data. Statistical significance was calculated with two-sided Wilcoxon signed rank test. Also, did web searching and coding of data by category (e.g., open or not). "Citation rate", I believe, is number of citations. This isn't explicitly stated.
2. Literature review
Is the literature review relevant to the population, setting, study objectives, and methodology used for the present research? The literature review includes several recent studies on the research impact of open access journals

and articles. Literature on "what is OA" is not discussed, although the "two roads" idea is mentioned right away. Conflation of studies relating to both OA archiving and OA publishing could be problematic, because she is unequivocally studying the former only. The literature review also tries to set up why the question is important and why authors seem to be adopting OA—tying this to impact. Overall, the lit review focuses more on what people have tried to learn from studies on OA rather than literature on citation impact or bibliometrics. The reason for the choice of the disciplines is explained by noting that each represents a different level of OA adoption. However, no bibliometric studies on the disciplines is cited (in other words, the expectation about OA adoption is not attributed to a particular reason).

Are included materials up to date and comprehensive?

The article itself is not current, but the literature is appropriately contemporary, including sources that are not formally published.

Do the authors appear to have evaluated the quality of the cited materials?

Some limitations of previous (or in-progress) studies on impact and OA are noted, although the methodologies are not reviewed in depth.

If a bibliographic database is used, have the authors considered published or unpublished evaluations of the database as a way to understand problems they may encounter?

The limits to ISI data as only "surrogates" to the idea of impact are noted (notes 18 and 19). However, the problems are taken to be minimal, since comparison is the key purpose, not absolute accuracy. One limit that is NOT mentioned, though, is that ISI doesn't necessarily include citations in open access publications—or any others that are not indexed there. This could change the citation data.

The limitations of Google are discussed at length, although no sources are cited with regard to information about Google's mechanics of indexing. Interestingly, Google Scholar was unveiled in November 2004, two months after this study was published. It is quite possible that different rates of open access would be uncovered with this tool.

Does the literature review clearly support the need for the present research?

Especially on page 378, it is assumed that open access is a desirable thing. The study is situated in the context of librarians and their communication with researchers about OA. The logic becomes circular: OA is important because it gives the articles a higher impact; we need to find out if OA gives them a higher impact so that we know if they're important.

3. Data set: Acquisition and analysis

How were the data acquired, and is the process clearly explained? (i.e., from a bibliographic database, handsearching, etc.)

Handsearching or (or more accurately, handbrowsing) in a bibliographic database was used. For

political science and electrical engineering, articles were selected beginning at the mid-point of the time period and then extended equally forward and backward in time until the target sample population was reached. All articles for the time period in math and philosophy were used. It's not clear, though, what the "target sample size" was. Then article titles and number of citations were recorded for each article. The method isn't clear, though—were they exported to a citation management software? Or was it done by hand? Handsearching of titles was done in Google to determine open access availability.

Links to where OA articles were stored were checked and coded by hand.

What are the exclusion and inclusion criteria for the data?

Self-citations, citations from articles within the same journal issue, and citations from 2004 [the year of the article] were excluded. No articles from the sample itself were excluded—it seems that the selection was purely mechanical.

In the section in which the location of OA articles was checked, it is not mentioned if any articles from the random sample were excluded.

Do these criteria seem relevant to the research question(s)?

No reasoning for the exclusion of certain citations is offered. But presumably, such citations would not necessarily demonstrate the "impact" of the research, which is what the author is attempting to measure. Perhaps the 2004 citations are considered only to represent a partial calendar year and therefore not accurate?

Are retrieval methods described in sufficient detail to replicate the process?

The selection of sample articles seems to be clear, although the stopping point of the sample size is fuzzy. An endnote explains how philosophy journals were selected, since a different process was used. For the other 3 subjects, "10 top journals" were selected, but it's not clear if these were the top 10 according to a particular ranking criterion, or if it was just 10 from among the top. At any rate, an endnote lists the journal titles that were examined. It seems clear that full article titles were searched in Google, although it's not noted whether this was a phrase search or if other parameters of any kind were used. The definition of what the study considered to be "open access" was detailed, although it's problematic that this does NOT correspond with the basic definition of open access. In other words, is an article that has a draft freely available truly "open access" in the more generic sense?

Are the limitations of the data source(s) considered?

The limits to ISI data as only "surrogates" to the idea of impact are noted (notes 18 and 19). However, the problems are taken to be minimal, since comparison is the key purpose, not absolute accuracy. In the EBLIP review, Lewis (and even Antelman herself) notes that new methods of bibliometric analysis are arising to study new publishing and distribution trends. One limit that is NOT mentioned, though, is that ISI doesn't necessarily include citations in open access publications—or any others that are not indexed there.

The limitations of Google are discussed at length (p. 375), although no sources are cited with regard to information about Google's mechanics of indexing. Interestingly, Google Scholar was unveiled in November 2004, two months after this study was published. It is quite possible that higher rates of open access would be uncovered with this tool.

The issue of time delay is not addressed. It is conceivable that OA articles have been available longer (i.e., as pre-prints) than their locked counterparts. The reverse could be true as well, however. Either way, this potentially confounding factor isn't addressed, although the issue of changeable access is mentioned on p. 376. A true comparison can't be made if it's not certain that the articles have been available to researchers for the same amount of time.

Are sufficient examples (tables, figures, etc.) provided to help you understand the data handling processes?

Yes. Actual values, in addition to percentages, are provided. A box plot is used a supplement to give more meaning to the data.

If there are discrepancies or contradictions in the data, are they accounted for (i.e., missing or incomplete information)?

It is noted that the statistics on open access availability are incomplete, but it's probably not possible to know the "true" number anyway. There are no particular contradictions in the data. It is noted that rate of OA adoption does not necessarily correspond to OA impact.

Is the process for organizing the data logical and clearly explained?

Open versus not is pretty clear. The study is meant to test those two categories, so it makes sense. It is interesting to note, though, that aggregate data is not provided—i.e., are there patterns when articles are combined, regardless of discipline? Gven that the research question doesn't mention a comparison of disciplines, I was initially surprised that aggregate data is not provided. On the other hand, the large differences in sample size and OA adoption across the disciplines would render analysis completely useless.

If categories or themes are assigned as a way to group information about the data, are they …

A. appropriate and sufficient to respond to the research question(s)?

One table notes the location of the OA articles, which is interesting although not directly related to the research question. [Note that the author did publish a later study on authors' self-archiving practices using this same methodology.] The articles are also grouped according to discipline, which is interesting and provocative but perhaps not strictly necessary to the primary research question. Later literature suggests that Antelman was on the right track in looking at author behaviour and disciplinary differences.

B. derived from the data itself, or from prior research?
Categories are derived from the data itself.

C. validated by some means, such as double-checking by other trained researchers, with discrepancies identified and resolved?

No mention of this. No credits are given to research assistants, but it's hard to believe the principal researcher did all of the searching herself.

D. If so, are interrater reliability statistics provided?
No mention of this

E. defined and labeled using unambiguous terms?
Yes. In particular, open versus not open is clearly defined, if not matched to standard definitions. Some statistical notation may not be clear to casual readers.

4. Statistical analysis

Is the method chosen for statistical analysis appropriate for the question and the data?

The data is skewed across disciplines, so a nonparametric test was chosen. According to the Wikipedia article on the Wilcoxon signed-rank test, it is appropriate when the data cannot be expected to have a normal distribution (i.e., not a bell curve).

Is the analysis reproduceable?

Yes. If I knew how to do a Wilcoxon test, I could!

Are the parameters for statistical significance established and explained?

A non-parametric test is used.

Are outliers (anomalies in the findings) discussed in terms of cause and effect?

There are no particular anomalies noted. However, cause and effect is implied in attributing a higher citation rate to open access, particularly on p. 377. Presentation of the data does suggest this causation.

5. Findings, contribution, and generalizability

Does the study achieve its original objective(s)?

The article measures research impact—as definted by the surrogate of ISI citation rates in terms of correlation with open access availability. Larger questions are also addressed and raised in the article, but the data itself is provided to answer the core research question. Data supporting some of the other assertions is sometimes lacking, e.g., about reader behavior and author behaviour.

The author successfully answers the yes/no question of whether articles that are available as OA have a higher mean citation rate. However, the question of "greater research impact" is not (cannot?) be definitively answered if a broader definition is taken. A more useful study might look at the actual use of research studies beyond simply a count of citations. Antelamn also cannot provide a

"why" explanation to the patterns that she discovered, although a few suggestions are mentioned (and sometimes dismissed).

If the study builds on prior research, does the present study validate, refute, or add to the earlier findings?

The study does continue a trend of anecdotal perceptions and studies finding that open access availability is related to higher citation rates. It also adds further data to the study of article level impact. It does indeed add data to a rapidly changing publishing landscape. It does not clear up the question of causation and selection bias (i.e., articles that are higher quality are more likely to be posted online freely available). Since the article is from 2004, subsequent research, building on this article, would be more helpful to librarians.

If there are discrepancies or contradictions in the findings, are they discussed?

One discrepancy is noted on p. 378: the standard deviation for open access articles is higher for 3 of the disciplines, but not in philosophy. This is not discussed further, simply noted. It is also noted that rate of OA adoption does not necessarily correspond to OA impact (p. 377).

Can the study by generalized beyond the setting or data examined in this study?

Since the overall pattern is cross-disciplinary, it seems that the simple statement of "more likely to be cited" might apply in other disciplines. However, more data is needed to be certain. Also, individual disciplines have vastly different citation practices, so the more specific numbers definitely cannot be generalized. Medicine is a discipline that is strikingly absent, since it has very distinctive OA and citing practices. There is also an element of timing—the publishing landscape was changing and has changed since this study was done.

What information do you need to obtain locally to assist you in responding to the findings of this study?

A correlation to the location of the article would be useful—i.e., are the articles on authors' websites more likely to be cited than those in institutional repositories or discipline repositories? This is something that is addressed in Antelman's response to Davis' letter: author selection of which articles to post, and how and where they are posted, is significant.

A stronger test of causation would be awesome, but isn't possible in this study.

A two-year time period for bibliometric studies is standard per Garfield, but it could be interesting to go further back in time. Including a data source that indexes a wider variety of literature (e.g. Google Scholar) would also be interesting.

Can the findings be used in your setting?

Is this an article that we want to bring to the attention of our faculty when discussing open access? The age of the article limits its usefulness now, but it can still be immensely useful to open discussions with faculty by drawing attention to articles like this. Without causation proven, it can't

be used as a simplistic guarantee argument to encourage researchers to deposit OA, but it can still open conversations about how and why people share their research through OA methods. It can also raise awareness about

Further research, of course, is important—particularly studies of articles that are actually published in OA journals. Concrete examples in particular disciplines are also important. But overall, articles like this can help in forming a "vision" of what OA is and how it relates to research and scholarly communication.

Does the disciplinary difference tell us something about how we can better provide and promote OA publishing and our own repository?

Can the methods used in this study be used in your setting
The study could be replicable in our setting

Also potentially problematic: p. 375. "Any freely available free-text version (including drafts, preprints, and postprints) was available, the article was considered to be open access." But this doesn't match the **definition** at the beginning—the two roads. Does it matter whether they were intentionally OA? On pg. 374, she says the authors' intentions are not considered, but could that be relevant in defining OA?

Not sure about this conclusion: "the **greatest impact** of open access is with the most-cited articles" on page 378, first column. What does "greatest impact" mean? I think that perhaps since the standard deviation is higher for all but philosophy, it means there is more distance from the mean in the data, meaning that the more an article is cited, the more the open access availability increases the citation.