2013

Practical Applications of Do-It-Yourself Citation Analysis

Steve Black

seblack@colgate.edu

Recommended Citation

Black, Steve. (2013). "Practical applications of do-it-yourself citation analysis." *The Serials Librarian* 64, 285-298. doi:10.1080/0361526X.2013.760420

Practical Applications of
Do-It-Yourself Citation Analysis
Steve Black

Summary:

Measures of impact published by Elsevier and Thomson Reuters are useful for collection development, but the data is expensive and it reflects citations from all disciplines. Custom do-it-yourself (DIY) citation analyses allow one to create ranked lists of journals in specifically targeted sub-disciplines or areas of interdisciplinary study. A method for independently analyzing citations to create a ranked list of journals is described. Two methods for testing reliability of ranked lists are described, one employing Spearman's rho rank correlation, the other using coefficient of variation. Strengths and weaknesses of DIY citation analysis are discussed. Tips for conducting DIY citation analysis for publication are offered, and practical applications are summarized. An annotated bibliography of important literature on citation analysis is appended.

Rationale for Do-it-yourself citation analysis

The study of patterns and frequencies of citations is an objective, quantitative way to measure the impact of journals, authors, institutions, or nations. The reason librarians should consider independently conducting citation analyses is to help identify journals for supporting research, building collections, or submitting papers for publication. Published impact factors provide useful information, but the data from Scopus and the Web of Knowledge reflect citations to journals from all disciplines. Custom analysis is needed to determine the highest impact journals within a specific sub-discipline or interdisciplinary area of study.

Impact factors measure the relative frequency of citation to a journal. The basic formula is impact factor = cites to articles published in last 2 years ÷ number of articles published in last 2 years. Thomson Reuters' complete definition, a description of adjustments for self-citation, and a list of caveats is available at http://thomsonreuters.com/products_services/science/free/essays/impact_factor/. The validity of impact factors has attracted much attention, so a substantial body of literature addresses not only technical aspects of calculating impact but also the uses (and misuses) of impact factors. The annotated bibliography below provides key entry points into this literature. In brief, the primary critiques of relying on impact factors are:

- Data errors, both in the citations themselves and in how they are compiled by the citation indexes.
- The Matthew effect, coined by Robert Merton after the passage in Matthew (13:12), "for unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath." The more attention a work gets, the more attention it gets, skewing citations to papers that are notable for being noted, not necessarily because they're the highest quality or the most important.
- Impact vs. quality vs. importance. Impact is how often cited. Quality is the caliber of the work—elegance of method, thoroughness of literature review, quality of writing, etc. Importance is contribution to a line of inquiry, regardless of how narrow. For various reasons high quality and very important papers may not be cited much.
- Global vs. local impact. Citation in publications may poorly correlate with the needs of local patrons.

So although calculating impact provides useful, objective, quantified information, it is important to recognize that impact is but one of many relevant factors to consider when selecting journals.

Method for Do-it-yourself Citation Analysis

As noted above, impact factors such as those published in Thomson Reuters' *Journal Citation Reports* reflect citations from all the journals covered by their database, with no distinction between citations from journals of interest for collection development and citations from journals out of scope. A custom, local do-it-yourself citation analysis allows one to target exactly the set of journals that are important for local needs. The basic method is straightforward:

- Select a target population, e.g.
  - journals in a sub-discipline
  - group of researchers
  - subject(s) or keyword(s)
- Select a sample that represents the target population
- Compile works cited in the chosen sample
- Sort and count the works cited

The author used such an approach to analyze citations as reported in the article "Frequently Cited Journals in Forensic Psychology," *Psychological Reports*, v.110, no.1, (2012): 276-282. For this case of journals specific to forensic

psychology, four tools were used: WorldCat (to identify most widely held journals), PsycINFO via EBSCOhost (to identify works cited), RefWorks (to organize works cited), and Excel (to count and rank cited journals).

A few points regarding these tools are worth noting. Librarians' ability to use WorldCat holdings as a measure of importance may fade over time. As libraries shift subscriptions to packages rather than to individual titles, holdings in OCLC may have little correlation with how many libraries provide access to a journal. PsycINFO is particularly well suited to do-it-yourself citation analysis because all works cited are included whether the article is in full text or not, and are in a format that is easily exported. Some other bibliographic databases also include works cited for every record, but many do not. RefWorks is only one of many citation management tools that can be used for organizing citations. The author chose it for its price, interoperability with EBSCO databases, and ease of exporting data to Excel. The free program Zotero also works with EBSCO databases, but it's not designed to easily export data to Excel. Excel has COUNTIF and PivotTable functions that could be used to count times cited, but either will undercount if there are any variations in spelling, punctuation, or spacing. The author chose to avoid that potential pitfall by highlighting the titles and noting the count of highlighted rows.

The sample selected to represent the target population of forensic psychology journals reported in *Psychological Reports* was the 2007, 2008, and 2009 volumes of six widely held journals specific to forensic psychology: *American Journal of Forensic Psychology*, *Behavioral Sciences and the Law*, *Journal of Forensic Psychology Practice*, *Journal of Forensic Psychiatry and Psychology*, *Legal and Criminological Psychology*, and *Law and Human Behavior*. This sample yielded 19,565 citations, of which 16,518 were citations to serials. For this NASIG presentation, the author gathered the works cited to articles in the 2011 volume of *Law and Human Behavior*, a sample of *n*=1,378.

The process of gathering and sorting the data to create a list ranked by times cited is as follows:

1. Using PsycINFO via EBSCOhost, retrieve records for articles in the source journal, one issue at a time.
2. Add cited references to a folder.
3. Export the references to RefWorks. RefWorks creates a temporary folder of the downloads.
4. Create a folder for each issue's works cited. (One folder for each volume also works. Each issue was kept discrete in this case to allow testing for reliability at the issue level).
5. Move the downloaded cites to the appropriate folder.

6. Export references in tab delimited format.
7. Save as a text document.
8. Open the text file in Excel.
9. Use the column "Periodical Full" to sort journal titles alphabetically.
10. Review the sorted list to correct discrepancies, e.g. convert J Psych to Journal of Psychology.
11. Count citation frequency by highlighting each group and noting the count shown at lower left. Alternately, use the COUNTIF or PivotTable functions. The author found it just as quick to highlight and note the count, which also helps avoid missing titles due to minor variations in spelling and punctuation. To save time recording titles with very low counts that will not end up on a ranked list of top fifty or even one hundred titles, one may skip titles with counts less than 0.5% of the sample.
12. Copy and paste the titles and frequency counts to a separate worksheet.
13. Sort the list on the frequency column by largest to smallest.

One now has a ranked list of the journals most frequently cited in the sampled source journal. As will be discussed below, robust citation analyses require a sample of multiple volumes of several journals. The final ranked list will be the aggregation of all sampled titles and volumes. While one could create one large folder for all works cited, it is better to have separate folders and worksheets for each volume, and create a separate worksheet to record the aggregated data. Keeping each volume discrete allows statistical testing of the reliability of the ranked lists.

Testing Rank Reliability with Spearman's Rho

Almost any size sample is capable of creating a ranked list, but how consistently does it reflect reality? In other words, how reliable is the list? Reliability may be tested in two ways. On the macro level, the variability of ranked lists may be tested. On the micro level, variability in individual journal's times cited may be tested. An appropriate test for the reliability of ranked lists is Spearman's rho rank correlation. The formula for Spearman's rho is

$$r_s = 1 - \frac{6 \, \Sigma \, D^2}{N(N^2-1)}$$

. D is the difference in ranks between two lists, and N is the number of ranked items in the lists. This is the simple version of Spearman's rho that assumes no ties in ranks. See the annotated bibliography under "Technical" for sources of the more complex formula. In practice the two formulas

return very similar results even with some ties between rankings. But if this method is used for a paper to be submitted to a journal with readers with a sophisticated understanding of statistics, it would be wise to employ the more complex formula for Spearman's rho.

The method enumerated above will typically yield ranked lists containing varying numbers of journals. A few adjustments must be made to calculate Spearman's rho rank correlation. First, the number of ranked items in each group must be equal, so choose a cutoff point. The rankings displayed in Figure 1 [insert Figure 1] show a case where the December volume had only 22 titles with 2 or more times cited. A cutoff of N=22 was thus chosen to compare an equal number of ranked journals. Second, handle ties in ranks by averaging, e.g. the two journals tied for 2 and 3 are each ranked 2.5. Finally, assign the bottom rank to titles that do not appear in the ranking. In the case shown in Figure 1, blanks will be ranked 22.

As can be seen in Figure 2 [insert Figure 2], correlations among rankings derived from single issues of *Law and Human Behavior* are weak and scattered. A general rule of thumb is that coefficients above .70 indicate strong correlation. None of the correlations reach that level, many show very weak correlations less than .30, and indeed quite a few have negative correlations indicating inverse relationships. Thus there is very low reliability of ranks at the single issue level, which for this case was groups of works cited of approximately $n=300$. One may conclude that a ranking created from a single issue is not reliable. However, note that the most cited journal was the same for all six issues. The consistency of *Law and Human Behavior's* top ranking is not simply due to self-citation. The author's more complete analysis for the *Psychologcial Reports* article showed it to be consistently the most cited title across the six source journals. So even with a sample too small to rank journals, a most-cited journal may be identified.

If one issue's works cited are insufficient to create a reliable ranking, what about one volume? Figure 3 [insert Figure 3] compares the ranks derived from works cited in one issue of *Law and Human Behavior* ($n=1,378$) with the ranks drawn from the three volumes of six journals used in the paper published in *Psychological Reports* (n=16,518). The cutoff for ranks is again chosen to be twenty-two, so N=22. The difference in ranks, D, is squared to make all values positive and to accentuate large differences in rank. The sum of differences in ranks squared D^2 is 684. The Spearman's rho formula thus works out as:

r_s=1-(6*684)/22*(222-1)

$r_s = 1 - (4104/10626)$

$r_s = 1- .37$

$r_s = .63$

A rank correlation of .63 is not quite enough to reach a strong correlation of .70, but it is close. Based on these tests, one may surmise a rough rule of thumb for reliability of ranking by sample size shown in Figure 4 [insert Figure 4].

Testing Rank Reliability with Coefficients of Variation

The alternate method for testing reliability of rankings is to take a micro view at variation in individual titles' rankings across samples. This is done by calculating coefficients of variation, which is the standard deviation divided by the mean. Examples are shown in Figure 5 [insert Figure 5]. One can readily see without calculating standard deviations and means that the counts for *Law and Human Behavior* are pretty consistent, and those of *Psychology, Public Policy and Law* bounce all over the place. But while that may be readily apparent, there are two benefits of calculating coefficients of variation (a process quite easily accomplished in Excel). The first is it allows one to generate a sorted list of the most consistently cited titles. In general, titles further down ranked lists have higher coefficients of variation. But knowing which titles are exceptions is good to know for collection development, since consistently cited titles with relatively low use might be good additions to a collection. The other advantage to calculating coefficients of variation is that one can average the variations to get an alternate measure of the reliability of ranked lists.

Summary of issues of reliability

The statistical analyses used in this case suggest that so long as the sampled journals validly represent the topic, a sample of
- $n<1,000$ works cited can indicate the top journal (if any)
- $n>1,000$ can generate a rough indication of leading journals
- $n>10,000$ can create a useful ranked list

However, even separate samples of $n>10,000$ will yield different ranks, especially further down the lists, so a journal's place in any one ranking must ALWAYS be taken as an approximation of its "true" ranking. Since rankings reflect the complex reasons researchers cite one thing or another, there can never be a static, definitive ranking for journals in any topic. Reliability tests the author has performed with samples of journals in communication disorders suggest that a third of movement in rankings over time is due to random variation.

Strengths of Do-it-yourself Citation Analysis

Spending the time and effort to conduct an independent citation analysis to produce a ranked list of journals has these advantages:

- Can target a precise area. Whatever the local need is, that can be what the sample is built around. A new major or minor, needs of a new research group, whatever it may be, the target is customized to local needs.
- Yields data not otherwise available.
- Objective, quantitative collection development tool. For practical or political reasons it can be very helpful to have an objective measure to balance against subjective judgments.
- Results may be of fairly broad interest. They're publishable! Following the method described here, including statistical analysis of reliability, will result in a paper with a solid chance of being accepted for publication.
- If the project is well organized, data can be gathered and sorted by research assistant(s). If one has the good fortune of having a research assistant, the work for this kind of citation analysis is well suited to delegation of tasks.
- Interdisciplinary--can team with faculty outside the library. A citation analysis project naturally bridges information science with one or more other disciplines.

Weaknesses of Do-it-yourself Citation Analysis

- Very time consuming to do well. Far and away the most important reason not to tackle a project, it requires many, many hours of rather tedious work.
- A really robust ranking requires $n>20,000$, and even that's not definitive. No matter how many hours are put into data collection and analysis, a final, perfect list is not obtainable. This is due to the nature of what is being studied—what is cited is a moving target.
- Works cited may not be readily available. Few databases make it as easy as PsycINFO to gather and download works cited. Experiment with gathering citations in the area of interest before making any commitments.
- It's recreating the (very expensive) wheel. One is much better off working with the data in Scopus or Web of Knowledge if is available.

Tips for publication

- Don't be too parochial or narrow—editors must think their readers will be interested. If it's offered as a major somewhere, it's probably broad enough.

- BUT choose something not already in *Journal Citation Reports.*
- Be very thorough with the literature review before gathering data. Definitely search relevant disciplinary databases as well as the library literature.
- Gather citations and test the sample before nailing down the method. Test the whole process to work out any kinks up front, especially if work is to be delegated.
- Pay especially careful attention to sample selection. This is the area most sensitive to criticism by editors and peer reviewers, and rightly so. Is the sample valid on its face? That is, will readers look at the sample and think it makes sense? Is it large enough? Does it represent an appropriate range of journals? Is coverage international—and does it need to be?
- Group data and run statistics by journal volume. Rank correlations can be done with smaller or larger groupings, but keeping data discrete by journal volume works well. Spearman's rho can be calculated to compare the aggregated rankings each time volumes are added. When correlation is very high after the addition of another volume, one can be confident that the sample is sufficiently large.
- Consider submitting to a journal in the topic area. Researchers within a sub-discipline or area of interdisciplinary study are often interested in ranked lists of journals, and those researchers are more likely to read the study if it is cited in databases familiar to them.

Tips for Collection Development

- Use citation analysis to assess impact of specialized journals. Journals with promising titles and scope of coverage may not have as much impact as expected. For instance, infrequently cited titles in this case of forensic psychology included *American Journal of Forensic Psychology* and *Journal of Psychiatry and the Law*.
- Don't be surprised to find surprises. Along with specialized titles that receive unexpectedly low citations, some journals with more general coverage receive surprisingly high ranks. The good news is that an existing collection may serve new majors and courses better than one might have anticipated.
- Judge new titles by other criteria. Citation analysis is by its nature biased to established journals. New launches must be considered by scope, reputation of publisher and editors, etc.
- Use ranked lists for assessment evidence. Comparison of a ranked list with holdings can be used as concrete evidence of how well the collection meets the needs of researchers in particular areas of interest.

Summary: Practical Applications of DIY Citation Analysis

Do-it-Yourself citation analysis is a scalable process with applications that include:

- Quickly determine the top journal on a topic. If one journal is clearly the most cited in an area, a small sample is capable of revealing it. Compile cites from 10-20 articles on the topic to discover if there is a definite most-cited title.
- Assess which journals a library should add to support a proposed new major, minor, or courses. The need here is to flag highly cited titles not currently in the collection. Compile cites from one or two volumes of two to four journals to discover high ranking titles not already available.
- Add objective analysis in cases where interested parties have difficulty choosing which journals to collect. If members of a department just can't decide, analyze publications and dissertations from the last decade or so to generate a ranked list of what has actually been used.
- Publish in a peer reviewed journal. If publication is required for tenure or promotion, a careful analysis of two or three volumes of four to six journals with reliability tests is likely to be accepted for publication. This assumes the topic area is of interest, the literature review is thorough, and the sample is valid.

Do-it-yourself citation analysis is time consuming and tedious, but can produce information that is very useful to librarians and researchers, both locally and globally.

Citation Analysis: A Selective Annotated Bibliography

The purpose of this bibliography is to suggest the most fruitful entry points into the substantial literature on using citation analysis to rank journals.

CLASSICS
S. C. Bradford, "Sources of Information on Specific Subjects," *Engineering* 137 (1934): 85-86.
    This study of the concentration of citations to literature in applied geophysics and lubrication is the original source of Bradford's Law of Distribution.

Eugene Garfield, "Citation Analysis as a Tool in Journal Evaluation," *Science* 178 (1972): 471-479.

Garfield explains and argues for the Institute for Scientific Information's *Science Citation Index*.

Robert K. Merton, *The Sociology of Science: Theoretical and Empirical Investigations* (Chicago: University of Chicago Press, 1973).
Includes the famous sociologist's description of how the Matthew Effect impacts scientific output.

OVERVIEWS
Linda C. Smith, "Citation analysis," *Library Trends* 30 (1981): 83-106.
One of several important articles in a special issue of *Library Trends* devoted to bibliometrics in the early years of the "serials crisis," when many libraries were first seriously confronted with having to decide which journals to cut.

Thomas E. Nisonger, "Chapter 5, The Application of Citation Analysis to Serials Collection Management," *Management of Serials in Libraries* (Englewood, CO: Libraries Unlimited, 1998): 121-156.
A well-organized presentation of major issues and a thorough bibliography.

Thomas E. Nisonger, "Journals in the Core Collection: Definition, Identification, and Applications," *Serials Librarian* 51, no. 3/4 (2007): 51-73.
Summarizes ten methods for creating lists of core journals and discusses applications of core lists.

CRITICAL ANALYSES
Michael H. MacRoberts and Barbara R. MacRoberts, "Problems of Citation Analysis: A Critical Review," *Journal of the American Society for Information Science* 40, no. 5 (1989): 342-349.
A thorough and well organized critique with specific emphasis on *Science Citation Index*.

Per O. Seglen, "Why the impact factor of journals should not be used for evaluating research," *British Medical Journal* 314 (1997): 498-502.
Concise summary of the problems associated with the use of journal impact factors.

Maurice B. Line, "Changes in Rank Lists of Serials Over Time: Interlending versus Citation Data," *College & Research Libraries* 46, no. 1 (1985): 77-79.

Robert N. Broadus, "A Proposed Method for Eliminating Titles from Periodical Subscription Lists," *College & Research Libraries* 46, no. 1 (1985): 30-35.
Maurice B. Line, "Use of Citation Data for Periodicals Control in Libraries: A Response to Broadus," *College & Research Libraries* 46, no. 1 (1985): 36-37.

> These 3 articles comprise a dialogue between Broadus and Line on the validity of citation data as a tool for collection development in an era of cost-cutting.

R.E. Rice, Christine L. Borgman, Diane Bednarski, and P.J. Hart, "Journal-to-Journal Citation Data: Issues of Validity and Reliability," *Scientometrics* 15, no. 3 (1989): 257-282.

> Reviews issues of validity and reliability, discusses causes of measurement errors, and concludes with suggestions for how to reduce measurement errors.

Ben R. Martin, "The Use of Multiple Indicators in the Assessment of Basic Research," *Scientometrics* 36, no. 3 (1996): 343-362.

> Defines quality, importance, and impact. Emphasizes the importance of respecting each.

Gordon and Breach Science v. American Institute of Physics and American Physical Society, http://barschall.stanford.edu.

> This web site contains a thorough and well organized treatment of the Gordon and Breach case against Henry H. Barschall and the publishers of his studies. Barschall used citation counts and subscription costs to create rankings of physics journals. Gordon & Breach titles were shown to be among the poorest values in physics, and they sued the publisher for false advertising. The court records and related documents do an excellent job of presenting the issues surrounding applications of citation analysis.

METHODS
Steve Black, "Using Citation Analysis to Pursue a Core Collection of Journals for Communication Disorders," *Library Resources & Technical Services* 45, no. 1 (2001): 3-9.

> Includes a basic method for do-it-yourself citation analysis.

Steve Black, "How Much do Core Journals Change over a Decade?" *Library Resources & Technical Services* 56, no.2 (2012): 80-93.

> Describes methods for correlating ranked lists over time.

Jeffery D. Kushkowski, Kristin H. Gerhard and Cynthia Dobson, "A Method for Building Core Journal Lists in Interdisciplinary Subject Areas," *Journal of Documentation* 54, no. 4 (1998): 477-88.
> Describes a Simple Index Method for ranking journals based on results of subject or keyword searches in relevant databases.

Daniela Rosenstreich and Ben Wooliscroft, "Measuring the impact of accounting journals using Google Scholar and the g-index," *British Accounting Review* 41 (2009): 227-239.
> Valuable for its treatment of Google Scholar, a table summarizing common criticisms of citation-based journal rankings, and comparisons of ranking methods. (Also a good example of how an important paper can be published in an unexpected place!)

Chris Piotrowski, "Top cited journals in forensic psychology: An analysis of the psychological literature, "*American Journal of Forensic Psychology* 30, no. 2 (2012): 29-37.
> An example of using keyword searches to rank journals. Piotrowski's method yields a very different ranked list from this author's list published in *Psychological Reports*.

TECHNICAL
Thomson Reuters, "The Thomson Reuters Impact Factor,"
http://thomsonreuters.com/products_services/science/free/essays/impact_factor.
> An overview by Eugene Garfield that includes the formula for calculating impact factor, the rationale for using it, caveats and cautions.

Stephen J. Bensman, "Probability Distributions in Library and Information Science: A Historical and Practitioner Viewpoint," *Journal of the American Society for Information Science and Technology* 51, no. 9 (2000): 816-833.
> Argues that parametric statistics based on Poisson distribution are incapable of accurately modeling patterns of journal citations.

Sidney Siegel, "Nonparametric Statistics," *American Statistician* 11, no. 3 (1957): 13-19.
> An authoritative, readable description of when and why to use various statistical methods including Spearman's rho.

Maurice Kendall and Jean Gibbons, *Rank Correlation Methods. 5th ed.* (New York: Oxford University Press, 1990).

Complete descriptions, formulas, and proofs of Spearman's rho and Kendall's tau (a calculation based simply on whether items go up or down in rank, disregarding the degree of change).