



Evidence Summary

External and Internal Citation Analyses Can Provide Insight into Serial/Monograph Ratios when Refining Collection Development Strategies in Selected STEM Disciplines

A Review of:

Kelly, M. (2015). Citation patterns of engineering, statistics, and computer science researchers: An internal and external citation analysis across multiple engineering subfields. *College and Research Libraries*, 76(7), 859-882. <http://doi.org/10.5860/crl.76.7.859>

Reviewed by:

Stephanie Krueger
Head, Office of Specialized Academic Services
Czech National Library of Technology
Prague, Czech Republic
Email: stephanie.krueger@techlib.cz

Received: 30 June 2016

Accepted: 19 Oct. 2016

© 2016 Krueger. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

Abstract

Objective – To determine internal and external citation analysis methods and their potential applicability to the refinement of collection development strategies at both the institutional and cross-institutional levels for selected science, technology, engineering, and mathematics (STEM) subfields.

Design – Multidimensional citation analysis; specifically, analysis of citations from 1) key scholarly journals in selected STEM subfields (external analysis) compared to those from 2) local doctoral dissertations in similar subfields (internal analysis).

Setting – Medium-sized, STEM-dominant public research university in the United States of America.

Subjects – Two citation datasets: 1) 14,149 external citations from 16 journals (i.e., 2 journals per subfield; citations from 2012 volumes) representing bioengineering, civil engineering, computer science (CS), electrical engineering, environmental engineering, operations research, statistics (STAT), and systems engineering; and 2) 8,494 internal citations from 99 doctoral dissertations (18-22 per subfield) published between 2008-2012 from CS, electrical and computer engineering (ECE), and applied information technology (AIT) and published between 2005-2012 for systems engineering and operations research (SEOR) and STAT.

Methods – Citations, including titles and publication dates, were harvested from source materials and stored in Excel and then manually categorized according to format (book, book chapter, journal, conference proceeding, website, and several others). To analyze citations, percentages of occurrence by subfield were calculated for variables including format, age (years since date cited), journal distribution, and the frequency at which a journal was cited. Top journals for selected subfields were identified based on the percentages of authors citing them in each dataset and, for interdisciplinary journals, according to how often citations for them appeared in subfield groups.

Main Results – For each subfield group, distinct patterns emerged for both internal and external analysis in terms of format, currency, and preferred journals. Regarding format of material cited, journals were dominant for external citations and ranged between 40% of citations (CS) to 94% (bioengineering) of formats cited. Formats were more distributed for internal citations, with ECE, SEOR, and STAT exhibiting journal dominance (61%, 30%, and 59% of citations, respectively) and conference proceedings dominant in CS (43%) and AIT (30%). Regarding currency, almost all cited items (>98% for external citations and 96% for internal citations) were published within the last 50 years, with electrical engineering showing the highest percentage of materials cited within the past five years for external citations (47%). For internal citations, applied information technology illustrated the most use of materials in the five-year timeframe (46%). Top journals for each subfield in which only external data were analyzed include *Journal of Biomechanics* (bioengineering 54%), *Engineering Structures* (civil engineering 47%), *Water Research* (environmental engineering 60%). For CS and AIT, the top journal was *Communications of the ACM* (external CS citations 29%; internal CS 32%; internal AIT 36%). For electrical engineering, the top journals were *Electronics Letters* (21% external citations) and *Proceedings of the IEEE* (50% internal citations). SEOR was broken into three categories (systems engineering, SEOR, and operations research),

with *Systems Engineering* being the top journal according to external citations for the subfield of the same name (48%) and *Air Traffic Control Quality* as the leading SEOR journal (25% internal citations only). *Management Science* (77% external citations only) was the top journal for operations research. Top STAT journals were *Annals of Statistics* (96% internal citations) and *Journal of the American Statistical Association* (60%). *Science* was the top interdisciplinary journal for external citations (10%) and *IEEE: Transactions on Pattern Analysis and Machine Intelligence* for internal citations (13%).

Conclusion – An approach to citation analysis integrating both internal and external components is useful for institutions aiming to develop balanced STEM collections as well as for collection assessment and budgeting purposes and enables adjustment of serial/monograph ratios to create custom local serial/monograph ratio “blends.” In this institution’s case, internal data suggested a 59:41 serial/monograph ratios versus an external data ratio of 75:25, which indicated that a blended ratio of 67:33 might be appropriate for this institution based on an average of both ratios. In the future, cross-institutional collaboration for external analyses would make it easier for institutions to focus on internal analyses in order to develop appropriate local serial/monograph ratio blends.

Commentary

Citation analysis, considered a branch of bibliometrics (Hoffmann & Doucette, 2012), has been used in a variety of settings and across disparate populations in an attempt to describe how users interact with resources, making key assumptions in terms of validity that citations represent accurate snapshots of resource use in time and are of high quality (Beile, Boote, & Killingsworth, 2004). As Kelly notes in her literature review, many prior citation analysis studies have attempted to apply research findings to inform collection development, but they have used citation sets (i.e., datasets) that are 1) too narrow for use across institutions or disciplines, or 2) too

general to be applicable to individual institutional settings. Kelly, by including both external (global) and internal (local) datasets, attempts to overcome such limitations and to point the way toward future studies that might be comparable, reproducible, and therefore more broadly valid – all goals which prior studies have failed to achieve (Hoffmann & Doucette, 2012).

While failing to provide a methodological “holy grail” for reasons regarding sampling outlined below, Kelly’s study does follow guidelines developed by Hoffmann and Doucette (2012) for citation analysis studies: the author clearly describes the rationale for her study as well as the two samples (i.e., datasets) under investigation. She describes the specific steps undertaken to conduct her analysis, enabling reproducibility, and offers straightforward presentation of research results via analysis of variables for well-defined subfields. The presentation of variables includes comparisons between external and internal datasets, the former of which might be re-used and therefore applicable in future studies as a kind of control against which internal citations from other institutions, source types, or disciplines could be compared. Reproducibility could have been enhanced with a deeper description of how, for external citations, the varying impact indicators for Thomson Reuters Web of Knowledge, ISI Journal Citation Index, and SciMago Journals and Country Rank were reconciled with one another in the creation of the journal source lists.

One crucial way in which Kelly’s approach could be improved in relation to the Hoffmann and Doucette methodological criteria would be by providing explanations for why the datasets selected could be considered representative samples. In this study, the target thresholds of 1,500 external citations per subfield and 1,200 internal citations per dissertation subfield appear to have been arbitrarily selected; while they might have been chosen as saturation points (Hoffmann & Doucette, 2012), this is not

explicitly stated. And though Kelly notes dissertation citations were selected at random, there is no description of the randomization process.

Since Kelly identifies the importance of conference papers for some disciplines (CS and electrical engineering, ECE for both external and internal citations, and AIT for internal citations), future studies focusing on these disciplines might potentially be enriched with a conference paper dataset (or datasets), in which citations from conference proceedings – categorized into serial or monograph format – would be additionally analyzed and included in blended serial/monograph ratios.

In terms of broader significance, the external component of this study provides libraries unable to conduct their own studies with ammunition for justifying the purchase or retention of key English language subscriptions in selected STEM subfields. For libraries interested in conducting their own similar studies, this article provides them with a roadmap, although the process described is labor intensive and might be streamlined with automated citation harvesting and management of citations in database form instead of spreadsheets.

References

- Beile, P. M., Boote, D. N., & Killingsworth, E. K. (2004). A microscope or a mirror? A question of study validity regarding the use of dissertation citation analysis for evaluating research collections. *The Journal of Academic Librarianship* 30(5), 347-353.
<http://doi.org/10.1016/j.acalib.2004.06.001>
- Hoffmann, K. and Doucette, L. (2012). A review of citation analysis methodologies for collection management. *College & Research Libraries* 73(4), 321-335.
<http://doi.org/10.5860/crl-254>