



*Evidence Summary*

**Weak Correlation Between Circulation and Citation Numbers Suggests that both Data Points should be Considered when Deselecting Print Monographs**

**A Review of:**

White, B. (2017). Citations and circulation counts: Data sources for monograph deselection in research library collections. *College & Research Libraries*, 78(1), 53 – 65. <https://doi.org/10.5860/crl.78.1.53>

**Reviewed by:**

Melissa Goertzen

Consultant and Information Manager

Halifax, Nova Scotia, Canada

Email: [goertzen.melissa@gmail.com](mailto:goertzen.melissa@gmail.com)

**Received:** 11 June 2019

**Accepted:** 23 Oct. 2019

© 2019 Goertzen. 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.

DOI: 10.18438/ebliip29606

---

**Abstract**

**Objective** – To facilitate evidence-based deselection of print monographs, this study examines to what extent there are correlations between circulation data (past and future usage) and between the borrowing and citation of print monographs.

**Design** – Collections assessment project that used a variety of data sources and techniques, including Spearman's rank correlation coefficient, statistical analysis, and the analysis of circulation data, last-use dates, and citation data.

**Setting** – An academic library in New Zealand.

**Subjects** – Two ranges of books were chosen for the study: 591 (Specific Topics in Zoology) and 324 (The Political Process). From these ranges, monographs published prior to 2001 were selected as the study sample.

**Methods** – This project relied on two data sources: circulation data from the Library's ILS and citation data from Scopus. All data was downloaded to an Excel spreadsheet in preparation for analysis. The researcher examined call numbers, authors and editors, titles and subtitles, publication dates, circulation counts, dates of last check-in, total number of citations, number of citations from publications released in 2010 and on, and number of citations from institution-affiliated documents. Renewal data was omitted, as it

did not provide evidence of additional instances of use.

Where multiple copies of a specific title appeared in the data set, the researcher totalled all circulations and recorded the most recent check-in date. The researcher found that some titles in the study sample were generic and it was impossible to determine if citation data from Scopus linked to the monograph in the library collection. These titles were eliminated from the study.

Once data collection was complete, the researcher calculated two additional data elements: the number of months since the last check-in date and the number of citations from items published before 2010. Data in the Excel spreadsheet was analyzed using Spearman's rank correlation coefficient to determine the relationship between past and future usage and between circulation and citation data.

**Main Results** – Findings indicated that circulation and citation data are highly skewed. Many monographs in the study sample had never been borrowed and had few citations, while a small number of “celebrity titles” were borrowed or cited at a much higher rate than other monographs in the same classification.

Further, results indicated that historic circulation numbers are imperfect predictors of future probability that a book will be borrowed. When taking a high-level view of the collection, highly circulated books tend to be borrowed more often than average. However, when examining monographs at the title level, high circulation is more of a probability instead of a robust indicator.

An investigation of whether historic citation counts serve as an indicator of future citation followed previously established trends: monographs not heavily cited in the past are less likely to be cited in the future. Findings also found a weak correlation between local-institution monograph citation counts and total citation counts.

Finally, the results demonstrated a weak correlation between circulation and citation data. As a group, well-cited books are borrowed more often than others, but at the individual title level, the effect is too random for either data set to predict the other in a reliable way. As such, circulation data and citation data can not be used as a proxy for each other.

**Conclusion** – Neither circulation nor citation data can stand as full proxies of the value of a title. However, both provide information that reflects the status of a title within the scholarly community. In this environment, citation data should be considered equally with circulation figures. Both data points measure different phenomena and the weak correlation between them suggests that both are required to inform decisions about deselecting print monographs.

### Commentary

The goal of collection development activities is to build collections that meet users' information needs. Studies conducted over the past several decades indicate that usage statistics support decisions to deselect print monographs from the collection; evidence suggests that past use is the best predictor of future use (Dinkins, 2003). Citation data is not typically factored into these discussions, as it is seen as a metric that guides decisions about journals as opposed to monographs. However, some argue that citation data can be used to supplement measures like faculty input and frequency of use (Burdick, 1989). This paper investigates the relationship between circulation and citation data, and how both metrics inform decisions to deselect print monographs.

When evaluated using the “Evaluation Tool for Bibliometric Studies,” strengths and limitations of this study emerge (Perryman, 2009). The strength of the piece is its organization and detailed descriptions of the methodology, data collection activities, and study results. The author identifies all data points used to analyze the past and future use of print monographs, describes the search scripts crafted to pull citation data from

Scopus, and lists correlations between data points. Based on this discussion, information professionals at other institutions could replicate the study and compare results against those outlined in this paper. Essentially, the author provides a roadmap for other librarians wishing to examine how circulation and citation data inform deselection decisions.

Limitations of this study include incomplete data sharing. The author stated that the distribution of both circulation counts and citations are highly skewed, and that collecting citation data was error-prone and incomplete. It would have been beneficial if the author provided the numbers and percentages of non-circulated and non-cited books, along with highly circulated and highly cited books. This would give a full picture of the results and allow readers to judge the validity of findings. Also, while the author explained the search strategy to identify citation data for monographs in the text, it would be more helpful to provide the exact search string for other librarians wishing to replicate the study. Finally, in the last paragraph of the paper, the author mentions a study by Kousha and Thelwall (2014) that utilized APIs to harvest citation data and standardize title-level citation metrics. It would have been interesting to see a similar method used in this study, as it provides an efficient and modern means to assess collections of print monographs.

Overall, this study provides value to librarians working in the area of collection development and monograph acquisitions. The paper presents a low-cost and sustainable methodology that supports decisions to deselect print monographs based on readily available data.

## References

- Burdick, A. J. (1989). Science Citation Index data as a safety net for basic science books considered for weeding. *Library Resources & Technical Services*, 33(4), 367–73.
- Dinkins, D. (2003). Circulation as assessment: Collection development policies evaluated in terms of circulation at a small academic library. *College & Research Libraries*, 64(1), 46–53.  
<https://doi.org/10.5860/crl.64.1.46>
- Kousha, K. & Thelwall, M. (2014). An automatic method for extracting citations from Google Books. *Journal of the American Society for Information Science and Technology*, 66(2), 309–320.  
<https://doi.org/10.1002/asi.23170>
- Perryman, C. (2009). *Evaluation Tool for Bibliometric Studies*. Retrieved 30 June 2019 from  
<http://libjournalclub.pbworks.com/f/Journal%20Club%20Jan%2020%202011.pdf>