

Thesis and Dissertation Citations as Indicators of Faculty Research Use of University Library Journal Collections

Louise S. Zipp

Citation analysis is a long-standing collection-evaluation tool often undertaken to investigate one aspect of library collection use. Citations from theses and dissertations are much more easily and comprehensively gathered than are citations voluntarily supplied by faculty. Using four studies in geology and biology, the Kendall coefficient of rank correlation tests the degree of association between journals most heavily cited by graduate students and those titles most heavily cited in faculty publications. Positive associations are confirmed in three data sets. Additional descriptive analysis shows that the 40 titles most heavily cited in theses and dissertations consistently contained about 70% of the top 40 titles cited by faculty, including most of the 12-15 top titles. If results are replicated, thesis and dissertation citations can be reliably used as a surrogate for faculty publication citations in evaluations of the research portion of library collection use.

As research journal prices continue to climb, selectors search for tools that will enable them to understand better the demand on library journal collections. Studies of circulation data and in-library use contribute to our understanding of what parts of a library collection are being used. The link between demand for research materials and use is most easily achieved by direct analysis of the scholarly communications process. Evaluation of the titles used in the process of discovery or problem identification yields valuable information. The titles in which researchers

publish and those titles that they cite in their works also serve to elucidate real and potential demand on library collections.

Citation analysis as a surrogate for use has served to identify journals for selection, de-selection, or off-site storage (Kelland and Young 1994). In research universities, faculty tend to be the most stable and vocal parties using research library materials. By using locally-derived data, analysis of journal titles cited by this known clientele has been valued as a means to measure demand (Crotteau 1991; Haas and Lee 1991; McCain and

LOUISE S. ZIPP is Head, Collection Development Department, and Principal Bibliographer for the Sciences, Iowa State University, Ames (lzipp@iastate.edu). The author is grateful to Katherine W. McCain and to Michael Mark Noga for permission to use their published data; and Carolyn Erwin for preparation of the Iowa data. Manuscript received March 14, 1996; accepted for publication May 23, 1996.

Bobick 1981; Noga, Derksen, and Haner 1994; Walcott 1994b; Wible 1990; Zipp 1995). In these studies, however, the faculty whose publications were examined represented small, discrete research groups. Dykeman (1994) measured the ability of the Georgia Tech library to support research in all scientific disciplines. Instead of direct solicitation of faculty or use of department-generated lists, she searched INSPEC to discover faculty publications. Swigger and Wilkes (1991) used data from the Institute for Scientific Information (ISI) to discover references cited by authors affiliated with Texas Women's University. In some studies where the target population is smaller and more discipline-specific, the ease of identifying faculty citations from ISI products still outweighed the limitations of database coverage (Greene 1993; Schmidt, Davis, and Jahr 1994; Williams 1990).

Graduate students conduct research as part of a program of study that results in an advanced degree. Walcott (1994a) observed that graduate students were the most active serials users in the Biology Library at State University of New York at Stony Brook. She analyzed thesis and dissertation citations to determine the most useful serial titles to that group, and she wondered why most citation studies ignore this crucial element of research library clientele. In fact, a few studies have examined this segment of research use (Chambers and Healey 1973; Herubel 1991; McCain and Bobick 1981; Noga, Derksen, and Haner 1994; Peritz and Sor 1990; Sylvia and Leshar 1995; Walcott 1991; Zipp 1995). Thomas (1993) used this approach to study use by the only research clientele at a teaching institution.

While graduate students are a transient population, their research interests, to some extent, reflect those of their faculty advisors. The product of graduate students' research is deposited in libraries and often indexed in general or thematic sources, such as Dissertation Abstracts International and GeoRef. Institution-specific studies of thesis and dissertation citations can use a full population or truly random sample of data, since the source documents can be identified and re-

trieved without involving the authors themselves.

Academic scientists in physics, chemistry, biology, mathematics, and geology depend primarily upon library subscriptions or their colleagues' personal subscriptions for retrieval of journal articles they subsequently cite (Hallmark 1994). Graduate students do not tend to have a network in place to make effective use of colleagues for discovery or retrieval of publications relevant to their research. Because they are more dependent upon library collections than are faculty, citation analysis of graduate student research may provide a more accurate snapshot of collection use.

In 1994, I undertook a comparative study to discover geology graduate student and faculty research journal use at the three state universities in Iowa (Zipp 1995). I was easily able to compile thesis and dissertation citations for a three-year period, but I encountered problems compiling even a representative sample of faculty citations. There were no institutional lists of faculty publications, and GeoRef, the primary index for geology, was known to be selective and delayed in its coverage. Given also that the earth sciences literature was not well represented in ISI products, I contacted faculty directly and requested information on their publications released from 1991-1993.

Schmidt, Davis, and Jahr (1994) suggested that faculty journal citations may be assumed also to represent graduate student use. McCain and Bobick (1981) had noted similarities among journal titles most cited in faculty publications and dissertations. They cautioned that data from dissertation citations, without considering those from faculty publications or Ph.D. qualifying briefs, were inadequate to determine the scope of research journal use. Noga, Derksen, and Haner (1994) observed that thesis and dissertation citations and faculty citations could not be used to predict each other. Given the value and viability of analysis of local data as a tool for refining library research journal collections, it is important to examine the strength of the relationship between these two measures of use. The purpose of this research is to determine to what degree graduate student research use of

journal collections can be expected to also represent faculty research use.

METHODOLOGY

To reduce the gaps in faculty citation coverage, described in Zipp (1995), I resolicited nonresponsive geology faculty in underrepresented subdisciplinary areas. While this follow-up was somewhat successful, it was also necessary to search GeoRef for publications by faculty who never responded. Some coverage was eventually achieved for all subdisciplinary areas, except earth science education. 2,127 faculty journal citations were compiled, and 1,208 journal citations were found in theses and dissertations. The two sets of forty journal titles most frequently cited by faculty and graduate students were combined into a sample of 52 titles (table 1).

A test of rank correlation was applied to reveal any comparability of the relative value of titles to each user group. The Kendall coefficient of rank correlation, Kendall's τ , represents the net proportion of concordant pairs in a sample, from which the proportion of discordant pairs has been removed (Gibbons 1993). The test is appropriate for samples with numerous ties. The titles were ranked, and the test was applied as described in Sokal and Rohlf (1995), including the use of correction factors for ties.

Few recent citation studies compare graduate student and faculty citations, and none use the Kendall coefficient of rank correlation to test the strength of a relationship. To provide a context for interpretation of my findings, I applied the test to three sets of published data, two in the geological sciences and one in biology. Collectively, these results can establish a baseline for subsequent use of this test in citation analysis.

THE STANFORD AND UCLA STUDIES

Use studies were done at the UCLA Geology/Geophysics Library and the Stanford Branner Earth Sciences Library (Noga, Derksen, and Haner 1994). As part of the study, geology theses and disserta-

tions for 1990 and 1991 deposited at each university, as well as a few more to complete subject coverage, were analyzed for citations to journals and monographic series held at each respective library. Stanford theses and dissertations yielded 7,652 citations, and UCLA sources yielded 2,893 citations. At Stanford, bibliographies of faculty publications from 1991 were analyzed for citations to journals and monographic series. At UCLA, faculty publications were analyzed for both 1990 and 1991. The Stanford faculty sample contained 3,460 citations, and the UCLA sample included 1,402 citations. I analyzed the data from each institution separately because of the quantity and reliability of information. In both cases, the same methodology devised for analysis of the Iowa citations was also used. The combined Stanford sample of the forty most heavily cited titles by both groups actually included 52 titles. The combined UCLA sample contained 56 titles.

THE TEMPLE STUDY

An earlier study in the biological sciences provided 1,793 journal citations from faculty publications and 632 citations from dissertations (McCain and Bobick 1981). The authors had analyzed Ph.D. qualifying briefs, dissertations, and scholarly publications of full-time faculty in the Temple University Department of Biology for 1975-1977. From the authors' table of the most highly cited titles, I extracted a combined sample of dissertation and faculty citations using the same methodology as employed with the Iowa analysis. The combined sample consisted of 60 titles.

RESULTS

For the Iowa data, a Kendall's τ of 0.3415654 was calculated. The range of possible values for τ is -1.0 to +1.0, representing a perfect negative relationship and a perfect positive relationship, respectively; a τ -value of 0 signifies no relationship. According to Gibbons (1993), a τ -value is not comparable to the more commonly used r_s -value of the Pear-

TABLE 1
 JOURNALS MOST HEAVILY CITED BY IOWA GEOLOGY AUTHORS

Journal Title	Thesis and Dissertation Citations	Faculty Citations
<i>Geological Society of America Bulletin</i>	80	121
<i>Journal of Paleontology</i>	62	133
<i>Journal of Sedimentary Petrology</i>	62	89
<i>American Association of Petroleum Geologists Bulletin</i>	56	65
<i>Geochimica et Cosmochimica Acta</i>	48	64
<i>Economic Geology</i>	41	28
<i>Journal of Geology</i>	38	22
<i>Soil Science Society of America Journal</i>	38	11
<i>Abstracts with Programs—Geological Society of America</i>	36	116
<i>Ground Water</i>	33	23
<i>Journal of Geophysical Research: JGR</i>	32	67
<i>Journal of the Iowa Academy of Science: JIAS</i>	32	59
<i>Contributions to Mineralogy and Petrology</i>	30	60
<i>Geology</i>	30	88
<i>Science</i>	30	49
<i>American Mineralogist</i>	24	47
<i>Geoderma</i>	24	0
<i>Earth and Planetary Science Letters</i>	21	66
<i>Journal of Hydrology</i>	21	15
<i>Nature</i>	20	46
<i>Paleobiology</i>	20	14
<i>Water Resources Research</i>	19	73
<i>Canadian Journal of Earth Sciences</i>	18	26
<i>Soil Science</i>	18	3
<i>Bulletin of the Centre of Excellence in Geology</i>	15	0
<i>Quaternary Research</i>	13	29
<i>Journal of the Geological Society, London</i>	12	27
<i>Lethaia</i>	12	0

TABLE 1 (continued)

Journal Title	Thesis and Dissertation Citations	Faculty Citations
<i>Sedimentology</i>	12	28
<i>Journal of Ecology</i>	11	3
<i>Bulletin of Marine Science</i>	10	7
<i>Canadian Geotechnical Journal</i>	10	4
<i>Eos</i>	10	28
<i>Marine Ecology Progress Series</i>	10	0
<i>Marine Biology</i>	9	4
<i>Philosophical Transactions of the Royal Society of London</i>	8	9
<i>Restoration and Management Notes</i>	8	0
<i>Tectonophysics</i>	8	11
<i>Coral Reefs</i>	7	4
<i>Tectonics</i>	7	20
<i>Journal of Metamorphic Geology</i>	6	10
<i>Chemical Geology</i>	5	15
<i>Journal of Petrology</i>	5	32
<i>Ecology</i>	4	14
<i>Geological Magazine</i>	4	16
<i>Precambrian Research</i>	4	17
<i>Canadian Mineralogist</i>	3	10
<i>Journal of Environmental Quality</i>	3	16
<i>American Journal of Science</i>	0	45
<i>Geophysical Research Letters</i>	0	14
<i>Journal of Structural Geology</i>	0	11
<i>Mineralogical Magazine</i>	0	19

son product-moment correlation coefficient, whose value tends to fall farther from zero.

Because the τ distribution approaches normality with sample sizes greater than 30, a z-statistic of 3.57 was calculated (Sokal and Rohlf 1995). Using a one-tailed test, the probability of finding a τ of 0.3415654 is 0.0002, which is significant at a 5% confidence level. Thus, the null hypothesis of no relationship between the variables can be rejected. The alternate hypothesis of a positive relationship cannot be ruled out (table 2).

Kendall's τ for the Stanford data is 0.2250267, and the P-value is 0.0091, supporting a positive relationship between the variables, although the relationship is weaker than in the Iowa case (table 2). Kendall's τ for the UCLA data is 0.1080369, which, although positive, did not lead to a P-value large enough to reject the null hypothesis (table 2). The Temple analysis yielded a Kendall's τ of 0.3425445, which is high enough to infer a positive relationship between the two variables (table 2).

While the range of variability in these

TABLE 2
RESULTS OF TEST FOR ASSOCIATION

Study	Subject	Sample Size	τ -Value	P-Value
Iowa	geology	52	0.3415654	0.0002
Stanford (Noga, Derksen, and Haner 1994)	geology	52	0.2250267	0.0091
UCLA (Noga, Derksen, and Haner 1994)	geology	56	0.1080369	0.119
Temple (McCain and Bobick 1981)	biology	60	0.3425445	0.00003

findings was unexpected, it is important to note that the Iowa and Temple studies only considered journals, whereas the Stanford and UCLA studies also incorporated monographic series, including open-file reports and map series. Although these formats were important to the goals of the original studies, they may have diluted the effect of journals in these samples.

Even though the test has confirmed the relationship that most selectors would have already assumed, the lack of comparable published data does not quite answer the original question. Doing so requires a more descriptive interpretation of the original ranked lists set into a context of possible uses.

In most research university libraries, marginal subscriptions were canceled long ago, and new subscriptions are not often placed without evidence of local demand. Document delivery, interlibrary loan, or non-library sources are expected to supply clientele with articles published in more specialized journals. Selectors are more likely to need lists of core journals, to ensure protection of those titles, usually for a designated clientele. To this end, how effectively does research use by graduate students predict research use by faculty?

The four data sets used for the statistical test were reexamined to provide a descriptive measure of the extent that one variable might predict the other (table 3).

The ten titles most frequently cited in theses and dissertations did not effectively predict the ten titles most heavily cited by faculty. Somewhat more surprising is the consistency with which the top 40 titles used by graduate students predicted the top 40 titles used in faculty publications. Amplifying the strength of this association is the fact that the 40 titles most heavily cited by graduate students included nearly all of the top 12-15 titles most cited by faculty.

CONCLUSION

Journal citations in theses and dissertations are better indicators of faculty use than has been previously assumed. A test for rank correlation showed that three of four investigations confirmed similar relative value for the same titles. To validate this conclusion and provide a basis for comparison, more data sets must be tested with a statistical measure of rank correlation. As the strength of this association is defined, the value of this approach for prediction must be further investigated. A less informative, but more striking, descriptive analysis showed that in all investigations about 70% of the faculty's 40 most cited titles were among the 40 cited most heavily by graduate students. This result also needs further examination.

Citation analysis remains a respected technique of collection evaluation. It requires time and diligence, coupled with

TABLE 3
RESULTS OF PREDICTION ANALYSIS

Study	% of Top 10 Faculty Titles Appearing in Top 10 Thesis/Dissertation Titles	% of Top 40 Faculty Titles Appearing in Top 40 Thesis/Dissertation Titles	% of Top 12-15 Faculty Titles Appearing in Top 40 Thesis/Dissertation Titles
Iowa	60	70	100 (of top 15)
Stanford	50	71	100 (of top 12)
UCLA	42	69	75 (of top 12)
Temple	70	70	93 (of top 15)

consistency and good data-handling skills. By choosing a method of data acquisition that avoids voluntary submission, researchers can actually gather a true population of citations. The most heavily cited journal titles in theses and dissertations can be used as a surrogate for the titles most heavily used by faculty in their publications. When appropriate sample sizes are considered, selectors can develop core lists of journals critical to local users and fully representative of the research portion of collection use.

WORKS CITED

- Chambers, George R., and James S. Healey. 1973. Journal citations in master's theses: One measurement of a journal collection. *Journal of the American Society for Information Science* 24: 397-401.
- Crotteau, Mark. 1991. Support for biological research by an academic library: A journal citation study. Master's thesis, School of Information and Library Science, University of North Carolina, Chapel Hill.
- Dykeman, Amy. 1994. Faculty citations: An approach to assessing the impact of diminishing resources on scientific research. *Library acquisitions: Practice & theory* 18, no. 2: 137-46.
- Gibbons, Jean Dickinson. 1993. *Nonparametric measures of association*. Sage University Paper Series on Quantitative Applications in the Social Sciences, no. 07-091. Newbury Park, Calif.: Sage.
- Greene, Robert J. 1993. Computer analysis of local citation information in collection management. *Collection management* 17, no. 4: 11-24.
- Haas, Stephanie C., and Kate Lee. 1991. Research journal usage by the Forestry faculty at the University of Florida, Gainesville. *Collection building* 11, no. 2: 23-25.
- Hallmark, Julie. 1994. Scientists' access and retrieval of references cited in their recent journal articles. *College & research libraries* 55: 199-209.
- Herubel, Jean-Pierre V. M. 1991. Philosophy dissertation: Bibliographies and citations in serials evaluation. *Serials librarian* 20, nos. 2/3: 65-73.
- Kelland, John Laurence, and Arthur P. Young. 1994. Citation as a form of library use. *Collection management* 19, nos. 1/2: 81-99.
- McCain, Katherine W., and James E. Bobick. 1981. Patterns of journal use in a departmental library: A citation analysis. *Journal of the American Society for Information Science* 32: 257-67.
- Noga, Michael Mark, Charlotte R. M. Derksen, and Barbara E. Haner. 1994. Characteristics of geoscience serial use by faculty and students. In *Finding and communicating geoscience information*, edited by Connie Wick, 61-97. Proceedings, vol. 24. N.p.: Geoscience Information Society.
- Peritz, Bluma C., and Dina Sor. 1990. The use of libraries by graduate students in psychology as indicated by citations. *Collection management* 12, nos. 3/4: 11-23.
- Schmidt, Diane, Elisabeth B. Davis, and Ruby Jahr. 1994. Biology journal use at an academic library: A comparison of use studies. *Serials review* 20, no. 2: 45-64.
- Sokal, Robert R., and F. James Rohlf. 1995. *Biometry: The principles and practice of statistics in biological research*. 3d ed. New York: W. H. Freeman.
- Swigger, Keith, and Adeline Wilkes. 1991. The

The State of the Art



...working smart

*Information Industry
leaders in providing
subscription services,
article delivery & library
automation software.*

- REMO
- ROSS
- Renewal Express
- Financial Planner
- BACKSERV/BACKMED

<http://www.readmore.com>

READMORE ACADEMIC SERVICES

700 Black Horse Pike, Suite 207
Blackwood, NJ 08012
Phone: 1-800-645-6595

READMORE IS A BLACKWELL COMPANY

use of citation data to evaluate serials subscriptions in an academic library. *Serials review* 17: 41-52.

Sylvia, Margaret, and Marcella Leshner. 1995. What journals do psychology graduate students need? A citation analysis of thesis references. *College & research libraries* 56: 313-18.

Thomas, Joy. 1993. Graduate student use of journals: A bibliometric study of psychology theses. *Behavioral & social sciences librarian* 12, no. 1: 1-7.

Walcott, Rosalind. 1991. Characteristics of citations in geoscience doctoral dissertations accepted at United States academic institutions 1981-1985. *Science & technology libraries* 12, no. 2: 5-16.

———. 1994a. Local citation studies—A shortcut to local knowledge. *Science & technology libraries* 14, no. 3: 1-14.

———. 1994b. Serials cited by Marine Sciences Research Center faculty, University at Stony Brook, 1986-1991. *Science & technology libraries* 14, no. 3: 15-33.

Wible, Joseph G. 1990. Comparative analysis of citation studies, swept use, and ISI's impact factors as tools for journal deselection. In *IAMSLIC at a crossroads: Proceedings of the 15th Annual Conference*, edited by Robert W. Burkhart and Joyce C. Burkhart, 109-16. N.p.: International Association of Marine Sciences Libraries and Information Centers.

Williams, Doris. 1990. Using core journals to justify subscriptions and services. In *IAMSLIC at a crossroads: Proceedings of the 15th Annual Conference*, edited by Robert W. Burkhart and Joyce C. Burkhart, 123-34. N.p.: International Association of Marine Sciences Libraries and Information Centers.

Zipp, Louise S. 1995. Identifying core geologic research journals: A model for inter-library cooperative collection development. In *Changing gateways: The impact of technology on geoscience information exchange*, edited by Barbara E. Haner and Jim O'Donnell, 59-65. Proceedings, vol. 25. N.p.: Geoscience Information Society.