

What 37000 citations can tell

Adriaan Swanepoel

Library and Information Services
Tshwane University of Technology
Pretoria, South Africa
e-mail: Swanepoelaj@tut.ac.za

Abstract: A longitudinal study at the Tshwane University of Technology (TUT) used citation analysis to analyze the reference lists of 480 master's and doctoral (M and D) theses and dissertations submitted at TUT between 2004 and 2007. The purpose was to determine what types of information sources M and D students at TUT use most, how the patterns of use differ across the seven faculties of the university, and to what extent the Library and Information Services (LIS) keeps or provides access to the journals that are mostly used by M and D students. More than 37 000 citations were analyzed over the four-year period. The study found several similarities but also some distinct differences in the use of information sources across the seven faculties of TUT. It also identified more than 60 different information sources used by M and D students. With regard to journal use, the study found that out of 3 641 different journals cited, most journals were only cited once over a period of four years. However, a small percentage of journals were highly and/or frequently cited.

Keywords: Citation analysis, theses and dissertations,

1. Introduction

1.1 Aims

The Tshwane University of Technology (TUT) started in 2004 with a longitudinal study to obtain information that would assist its Library and Information Services (LIS) in making sound collection development decisions. The aims of the study were to identify:

- Which types of information sources do master's and doctoral (M and D) students of different faculties in TUT use most and least, and to what extent does this change from year to year?
- Which journals do M and D students of different faculties in TUT use most and least, and to what extent does this change from year to year?
- To what extent does the TUT LIS keep or provide access to journals that are mostly used by M and D students?

1.2 Method

Library science has developed a number of methods for evaluating the use of information sources. Some of the more popular methods include a) studying circulation statistics and in-house use of information sources, b) surveying user opinions, c) analysing interlibrary loan patterns, d) doing shelf-availability studies, and e) analysing bibliographic citations. Although it is generally recognised that no one method alone provides a complete picture in assisting with collection development decisions, this study used citation analysis as the preferred research method to achieve its aims. This method, with its advantages and disadvantages, and how it was applied in this study, was described elsewhere by this author

(Swanepoel 2008). It is nevertheless necessary, for the sake of comprehensiveness, to repeat some of that information in this paper.

Citation analysis is known to be a low-cost method whereby researchers can gather and study citation data in an unobtrusive and non-invasive way. Citation analysis is also a flexible method: it can be applied in the assessment of data sources of a group of libraries, or a single collection or a library collection supplemented with external information sources. It is also flexible in terms of the size of samples, types of citation sources (whether a standard list or a specific collection in a library) as well as the manner of citation selection (Ching 2002). Furthermore, it can be used to focus on the type and number of information sources that researchers or library users use in a specific discipline or over a period of time. In this regard, the research of Peritz and Sor (1990) and Allen, Jacobs and Levy (2006) serve as examples.

Citation analysis is also a well-studied method in university library environments. Not only do citations play an important part in the scholarly communications process, but “citations and the composition of bibliographies reflect changes in the information-seeking behaviour of academics”, as well (Naudé, Rensleigh & Du Toit 2005). In this regard, theses and dissertations have proved to be particularly appealing to use for assessing library collections because they serve as a convenient source of in-house research. Furthermore, Zipp (1996) found that “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”. This is because the research interests of graduate students often reflect the research interests of their faculty advisers.

While many studies have reported on the uses and usefulness of citation analysis, some writers, including Loree (2007), Griscom (1983) and Ching (2002) have acknowledged the limitations of the method. Concerning citation analysis of students bibliographies, Sylvia (1998) highlighted several limitations, including the following: a) researchers are more likely to use information sources to which they have local access, b) researchers may add citations to increase a manuscript’s length and scholarship, c) manuscripts may include citations of marginal importance, d) researchers may not cite all works used to prepare a manuscript, e) handbooks and textbooks often do not receive citations because those sources are taken for granted by students.

Being aware of the limitations of citation analysis and acknowledging that the method is sensitive to the skill with which it is applied, it was nevertheless decided to use citation analysis as method for this research because it has been proved by many as a valid, reliable and practical tool for comparing a library’s holdings to an authoritative list for the purpose of evaluating the quality of a library collection.

1.3 Data collection

Data were collected from the reference lists of all (480) theses and dissertations submitted by M and D students and accepted by TUT from 2004 to 2007. No sample was taken. However, one thesis from the Faculty of Science was not included because its extreme number of citations (3 750) would have skewed results.

A library assistant photocopied the title pages and reference lists of all theses and dissertations and provided each data source with a sequential number. The assistant then captured the following data from the title pages on an Excel worksheet: a)

author, b) title, c) date, d) faculty, e) department at the faculty, f) language of the thesis, and g) whether it was a master's thesis or a doctoral dissertation. Subsequently, the researcher collected the following data: h) number of citations per information type, i) number of citations per thesis, j) number of theses per department, k) frequency of journal titles cited, and l) cited periodicals owned or provided access to by the library. All the data collected (a – l) was captured on Excel worksheets.

1.4 Data analysis

With regard to Aim 1: The different types of information sources used by M and D students in each of the seven faculties were counted and the numbers obtained per type were summarized per annum. The data were also sorted and ranked to identify which types of information sources M and D students in different faculties use most and least, and to what extent the pattern changes from year to year.

With regard to Aim 2: A list was made of all the different journal titles that were cited by M and D students in each faculty. The researcher then counted the number of times each title was cited per annum. This enabled him to determine which journals M and D students in different faculties use most and least, and to what extent this changes from year to year.

With regard to Aim 3: Each journal title cited by M and D students was compared against a list of journal titles that the LIS makes available either (a) by keeping those journals in stock or (b) by providing online access to full text versions of the journals. This process resulted in identifying to what extent the LIS keeps or provides access to the journals that are mostly used by M and D students.

1. Findings and discussion

2.1 Types and number of information sources cited

To put the findings of this study into context, it is necessary to present an overview of the types and numbers of information sources cited by the students under discussion. Although M & D students of TUT used 64 different types of information sources from 2004 to 2007, it is clear from Table 1 that they largely relied on books and journals for research purposes. Together, citations to books and journals add up to almost 70 percent of all citations. Furthermore, the top nine sources are typically those that are readily accessible in or through libraries. One should, however, not ignore the fact that several other information types play an important role in the research activities of M & D students, especially in-house documents, course material, technical publications, user manuals and newsletters /bulletins.

Citation analysis studies usually refer to less cited information sources as "other" information sources. Although this study also used the term "other" to group information resources that individually received less than one percent of all citations, it went a step further to actually identify those information sources and rank them according to the number of times they were cited (see footnote to Table 1). In doing so this study found that M and D students used more than 50 types of information sources that are generally regarded as "grey literature". This indicated that, although M and D students predominantly use books and journals, reference librarians and collection developers should take cognisance of the wide variety of so-called grey-literature that M and D students use.

Table 1: Types and number of information sources cited

Information sources	Total citations 2004 - 2007	Percentage
Books and chapters in books	14 617	39.11
Journals and magazines	11 503	30.77
Websites and electronic databases	3 740	10.00
Government publications	1 387	3.71
Conference proceedings and papers	1 158	3.10
Reports	956	2.56
Newspapers	676	1.81
Theses and dissertations	601	1.61
Personal communication	571	1.53
Other **	2 169	5.80
TOTAL	37 378	100.00

** Consisting of unidentified types (234), in-house documents (205), course material (200), technical publications (176), operating manuals and user manuals (169), newsletters and bulletins (158), videos (116), standards and test methods (102), policy documents (70), planning documents (65), manuscripts and drafts (61), working papers (48), event programmes (47), statistical data (46), discussion and position papers (45), software and computer files (38), comic books (33), fact sheets (33), speeches (29), surveys and questionnaires (29), patents (27), pamphlets and brochures (25), codes of conduct (19), radio and television programmes (16), briefing documents and position papers (16), case studies (16), specifications (14), media releases (13), trade literature (12), maps (11), photos (11), rules and regulations (11), prospectuses (9), circulars (8), staff lists (8), logbooks (7), company profiles (6), abstracts (5), catalogues (4), kits (4), proposals (4), exhibit captions (3), event invitations (2), microfilm (2), posters (2), graphics (1), sound recordings(1), calendars (1), drawings and diagrams (1), instrumental reviews (1), mathematical models (1), memorabilia (1), notices (1), simulation tools (1) and testimonials (1).

It should be noted that the number of citations to Websites was slightly higher than that reported in Table 1. In the process of classifying citations into different types of information sources, citations were first classified by content and then by form. For instance, a citation to a government publication on the Web was classified as a government publication and not as a Website. Only citations to Websites that were not clearly identified as either a journal on the Web or a government publication on the Web, etc., were classified as 'Websites'. However, Figure 1 presents a more complete picture of the number of citations to electronic sources; it shows the number of citations to all Websites regardless of content, as well as citations to electronic databases and CD-ROMs. Regardless of the way citations to Websites were classified, this study showed that M and D students of TUT still make far less use of electronic formats for research purposes than librarians and faculty members generally suspect. However, seeing that there was a slow but gradual increase of electronic formats over four years, the ratio between citations to paper formats versus electronic formats may seem significantly different over the next four years.

Figure 1: Citations to paper formats versus electronic formats

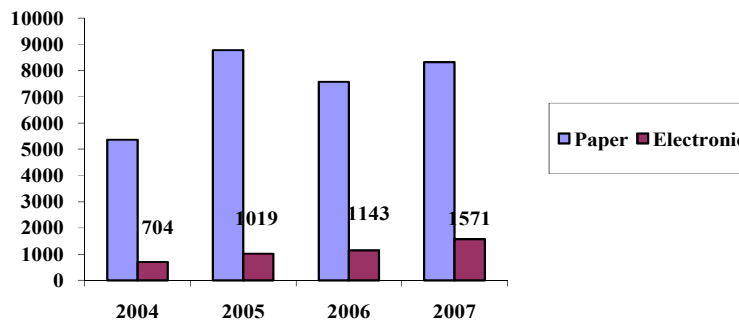


Figure 2 shows some similarities but also distinct differences in the use of information sources by M and D students in the so-called science, engineering and technology (SET) faculties, specifically regarding to books, journals and electronic formats. The Science Faculty clearly uses journals most, followed by books and electronic formats, conference papers and reports. Compared to all other faculties, this faculty uses journals most by far. The Faculty of Engineering also used journals primarily, followed by books, electronic formats, conference papers and reports, but its students use journals far less than students from the Faculty of Science. Engineering students also used electronic formats and conference papers more than science students. Students in the Faculty of Information Communication and Technology (ICT), on the other hand, first and foremost used Web sources; with books and journals respectively the second and third most used sources. It is also worth noting that ICT students were the only students that used Websites and electronic databases more than any other faculty's M & D students. Contrary to this is the relative low use that students of the Faculty of Science made of Websites and electronic databases. Not only was their use of electronic formats the lowest in the SET faculties but it was also the lowest in all faculties. The patterns showed in Figure 2 did not change much over the four years of investigation.

Figure 2: Types of information sources cited most by M and D students of science, engineering and technology (SET) faculties, 2004 - 2007

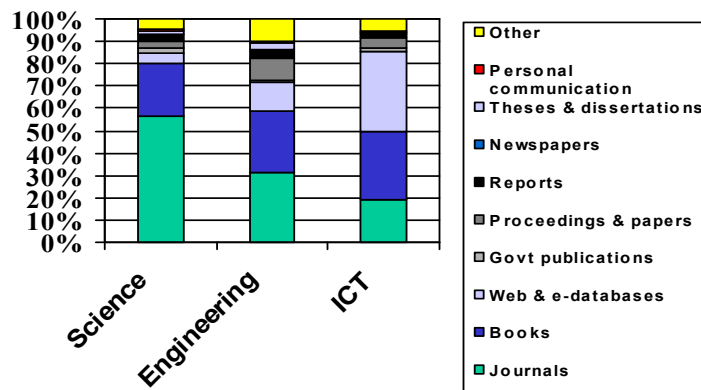
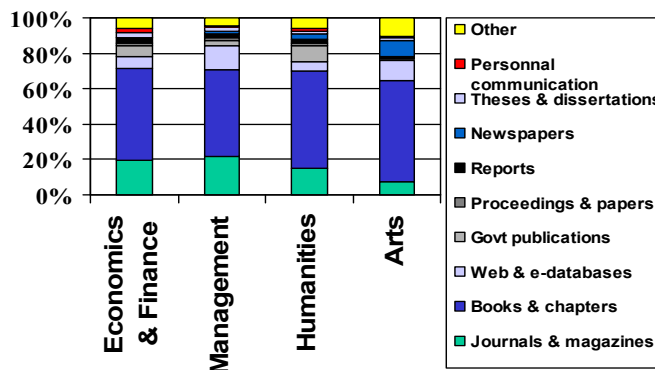


Figure 3 shows a different pattern from the one in Figure 2. M & D students in the social sciences and humanities (SSH) faculties clearly use books most. They also use noticeably fewer journals than students in the science, engineering and technology (SET) faculties. This is especially noticeable in the Faculty of the Arts. The use of journals in the Faculty of Arts ranked only fourth, next to books, electronic formats and newspapers. It is also clear from Figure 3 that the Faculties of Economic Sciences, Humanities, and Management Sciences cited government publications rather highly.

Figure 3: Types of information sources used by M and D students of social sciences and humanities (SSH) faculties, 2004 - 2007



2.2 Journals used by M and D students

One of the aims of this study was to determine which journals M and D students in different faculties use most and least, and to what extent this changes from year to year. The results produced a surprisingly large amount of data. Authors of the 480 theses and dissertations cited altogether 3 610 different journal titles in 2004 to 2007. The 3 610 journals received 11 533 citations. The average number of citations to individual journal titles was 7.49 and the average (mean) citations to journal articles were 23.93 over the four-year period.

2.2.1 Highly cited journals

Because of the large amount of data involved, this paper will display only a list of the top 10 most cited journals. A list of journals, across all seven faculties, that received 20 citations or more is available from the author.

Table 3: Ten most cited journals over a four-year period

Journal	2004	2005	2006	2007	Total
Harvard Business Review	9	73	3	24	109
International Journal of Pharmaceutics	20	52	25	5	102
Medicine and Science in Sports and Exercise	5	51	25	0	81
Journal of Analytical Atomic Spectrometry	25	38	7	8	78
Journal of Controlled Release	11	45	19	3	78
Journal of Applied Physiology	2	67	4	0	73
Spectrochimica Acta Part B	11	39	10	7	67
South African Journal of Higher Education	4	12	11	37	64
Human Reproduction	0	14	47	0	61
Pharmaceutical Research	14	27	19	0	60

The journals listed in Table 3 were not only highly cited but most of them were cited in all four years of investigation. However, when the researcher listed and ranked all journals cited in each faculty it was observed that the vast majority of journals had only been cited in one or two of the years concerned. For instance, of the 50 topmost cited journals across all seven faculties only 24 journals had been cited in the four years of investigation. This is probably an indication that the majority of journals that M and D students cited were used by a particular student only for a particular research topic. This tendency is more noticeable among some of the SSH faculties. For instance, of the 138 different journals that M and D students of the Faculty of Arts used during the four years of investigation, 133 (96.4%) journals were used only once. Five journals were used in two of the four years but no journals were used more than that. Contrary to this pattern, the M and D students in the SET faculties have shown a more consistent use of journals. Such consistent use will enable the LIS to identify a core list of journals that M and D

students in those faculties frequently consult.

2.2.2 Journals that received few citations

Table 4 presents the number of journals that received four or less citations. Note that the numbers in the columns do not overlap. In other words, if a journal was only cited once in 2004, it was not cited again in 2005 or in 2006 or in 2007

Table 4: Number of journals that received fewer than five citations per year

	2004	2005	2006	2007	Total
Journals cited only once	488	999	665	819	2 971
Journals cited twice	110	271	143	172	696
Journals cited three times	49	105	64	57	275
Journals cited four times	26	67	36	39	168

The majority of journals (2 971 or 82.3% of the total) were only cited once over a period of four years. Due to the fact that many journals were cited only once the median and the modus for each of the four years was only one (1) citation per journal title, notwithstanding that the mean was 8.28 in that period. The results in Tables 3 and 4 confirm results from other citation studies that a small number of journals generate the majority of journals cited, and *vice versa*.

2.2.3 The availability of highly cited journals in TUT libraries

A third aim of this study was to determine to what extent the LIS keeps or provides access to the journals that are mostly used by M and D students. The research found that of the 87 journal titles that received 20 citations or more, the LIS either subscribed to or provided online full-text access to a total of 58 titles. In addition it had back copies available in some of its libraries of a further eight (8) titles of which it recently cancelled the subscriptions.

3. Conclusion

From the findings of this study one can conclude that M and D students of TUT are far greater users of paper-based information sources than of electronic information sources. Over time there was a gradual increase in citations to electronic formats but not as much as one would have expected. Citations to electronic formats therefore had no real impact on the composition of reference lists.

With regard to paper-based sources, M and D students of TUT rely to a large extent on books and journals for research purposes. This, however, should not detract attention from the fact that several other information types play an important role in the research activities of M and D students, especially that of government publications, conference proceedings and papers, reports, newspapers, theses and dissertations and personal communication. Since many of the lesser used information sources were not part of the LIS collection one can therefore also conclude that rather than confining their research to what was available in TUT libraries, several M and D students made use of sources elsewhere. Even if lesser used information sources do not qualify to be included in the LIS collection, reference librarians of TUT should recognise the need for such sources and become

proficient in searching and retrieving them

This study revealed clearly distinguishable patterns between the information sources that M and D students in different faculties of TUT use most and least. This indicates that the LIS should consider separate collection development strategies for its seven faculties.

Across all faculties the majority of journals were only cited once over a period of four years. However, a small percentage of journals were highly cited and/or frequently cited. This confirms the outcome of other citation studies: A small number of journals generate the majority of journals cited. This confirmation can be used as a rationale for cancelling subscriptions of less-used journals or allocation of funds to faculties who consistently make high use of specific journals.

It is reassuring that TUT libraries are to a large extent able to provide most of the journals that are highly cited by M and D students. Although TUT libraries own only a few of the most cited journals they provide Web access to most of the others.

This study provided information that was previously unavailable to librarians and academic staff of TUT. It revealed what types of information sources M and D students used more frequently than others, how the patterns of use differ across subject disciplines and faculties, and it showed to what extent the LIS keeps or provides access to the journals that are mostly used by M and D students. The continued replication of this study will enable the LIS to determine even more reliable usage trends that will assist it to make informed decisions regarding which information resources to make available to M and D students, especially as Internet and electronic technologies evolve. A continuation of this research will also assist librarians and academics to better understand the research interest of M and D students at TUT.

References

- Allan, M., Jacobs, S. and Levy, J.R., (2006). Mapping the literature of nursing: 1996 – 2000. *Journal of the Medical Library Association*, Vol. 94, No.2, 206 – 220.
- Ching, J.T.Y., (2002). Collection evaluation through citation analysis techniques: a case study of the Ministry of Education, Singapore. *Library Review*, Vol. 51, No. 8, 398 – 405.
- Loree, S., (2007). Is citation analysis worth it: a comparison of local citation analysis, interlibrary loan records and usage statistics for collection development purposes in a special library. Master's Thesis, Chapel Hill, University of North Carolina. [Online]. Available WWW: <http://etd.ils.unc.edu/dspace/bitstream/1901/369/1/saraloree.pdf> (Accessed 8 January 2008).
- Griscom, R., (1983). Periodical use in a university music library: a citation analysis of thesis and dissertations submitted to the Indiana University School of Music from 1975 – 1980. *The Serials Librarian*, Vol. 7, No. 3, 35 – 52.
- Naudé, F., Rensleigh, C. & Du Toit, A.S.A., (2005). Analysis of the citation of Web-based information resources by UNISA academic researchers. *South African Journal of Information Management*, Vol. 7, No. 3, 1 - 14.

<http://www.sajim.co.za/peer99.7nr3.asp?print=1> (Accessed 22 March 2007).

- Peritz, B.C. and Sor, D., (1990). The use of libraries by graduate students in psychology as indicated by citations. *Collection Management* 12(3/4), 11 – 23.
- Swanepoel, A.J., (2008). Citation analysis of theses and dissertations submitted at the Tshwane University of Technology: 2004 – 2006. *South African Journal of Higher Education*, Vol. 22, No. 5, 1097 – 1113.
- Sylvia, M.J., (1998). Citation analysis as an unobtrusive method for journal collection evaluation using psychology student bibliographies. *Collection Building*, Vol. 17, No. 1, 20 – 28.
- Zipp, L.S., (1996). Theses and dissertation citations as indicators of faculty research use of university library journal collections. *Library Resources and Technical Services*, Vol. 40, No. 4, 335 – 342.