The Effect of the Web on Undergraduate Citation Behavior 1996–1999

Philip M. Davis
Albert R. Mann Library, Cornell University, Ithaca, NY 14853. E-mail: pmd8@cornell.edu

Suzanne A. Cohen
Martin P. Catherwood Library, School of Industrial and Labor Relations, Cornell University, Ithaca, NY 14853-3901. E-mail: sac29@cornell.edu

A citation analysis of undergraduate term papers in microeconomics revealed a significant decrease in the frequency of scholarly resources cited between 1996 and 1999. Book citations decreased from 30% to 19%, newspaper citations increased from 7% to 19%, and Web citations increased from 9% to 21%. Web citations checked in 2000 revealed that only 18% of URLs cited in 1996 led to the correct Internet document. For 1999 bibliographies, only 55% of URLs led to the correct document. The authors recommend (1) setting stricter guidelines for acceptable citations in course assignments; (2) creating and maintaining scholarly portals for authoritative Web sites with a commitment to long-term access; and (3) continuing to instruct students how to critically evaluate resources.

1. Introduction

Since the mid 1990s, the academic library has lost its control as the sole information resource provider on the college campus and now competes with a multiplicity of resources available over the Internet.

The wiring of the U.S. college campus has had a dramatic effect on how students access information. As a result, much of the research that was once done in libraries can now be done in computer labs or dorm rooms. According to statistics compiled for American College and Research Libraries (ACRL), reference statistics have declined 24% between 1995 and 1999. Circulation statistics have decreased 7% during this same period.

Anecdotal evidence from the reference desk informs us that students prefer electronic resources, lack the ability to distinguish credible academic sources from popular materials on the Internet, and have difficulty citing what they find.

David Rothenberg, an Associate Professor of Philosophy, claimed that he noticed a "disturbing decline in both the quality of the writing and the originality of the thoughts expressed [in student papers]," and attributed this to students doing their research on the World Wide Web (Rothenberg, 1998).

On the contrary, Kari Boyd McBride, lecturer in women's studies, and Ruth Dickstein, social sciences librarian at the University of Arizona, claim that they have not found that the Web has had an adverse affect on the quality of student writing (McBride & Dickstein, 1998).

In 1997, a study of 291 student bibliographies from ten colleges in Pennsylvania revealed little use of electronic media (Malone & Videon, 1997). There were difficulties, however, in accurately assessing student use of electronic resources because the researchers often had to speculate on whether a citation was retrieved online.

At Duke University, John Lubans surveyed 1,200 freshmen in 1997 on their use of the Internet (Lubans, 1998). Although the response rate was rather low (19.6%), the findings were quite valuable. Twenty percent of respondents believed that the Web has a positive influence on the quality of their written work, whereas 40% believe that the Web has made no difference. According to the study, more than half of the students believe that the Web is a time saver when looking for information, and has a modest influence on their grades.

More recently, D'Esposito and Gardner (1999) conducted an exploratory focus group of undergraduate perceptions of the Internet. According to the researchers, students are keenly aware of the importance of discerning reliable information from the Internet. The researchers reported that students felt that the Internet sites of highest quality and reliability were those produced by the government, educational institutions, and reputable businesses and corporations.

An unpublished report by Kate Majewski (1999) studied bibliographies submitted from 91 students in seven undergraduate classes at the SUNY College of Technology at

Delhi, NY and Hartwick College in Oneonta, NY. Students were also asked via a follow-up questionnaire as to their information seeking behavior. The author reported that students are still using traditional printed materials, but are neglecting traditional library locating tools, like library catalogs and journal indexes.

In summary, it is well accepted that the Internet has had an effect on the research behavior of undergraduates. What is not known is *how* this change in behavior has affected undergraduate scholarship. Although there is much anecdotal evidence on this topic, the literature lacks sufficient quantitative analyses to make substantial claims. What are also missing are studies that track the change in student research behavior over the period of World Wide Web adoption on college campuses.

1.1. Bibliometrics as a Tool for Understanding Research Behavior

Bibliometrics is a branch of library science concerned with the mathematical and statistical analysis of bibliographies. It has been used extensively for studying the publication record of research faculty and their departments, and also to understand how scholars communicate through their publications. There has been very little application of bibliometric research to undergraduate research papers. The principal cause of this problem is that research papers are returned to the student after grading, leaving no repository of their work.

Although survey research has been used to understand how students search for and use information, this form of research relies on the subject to objectively recall his/her own past behavior. Studying the bibliographies of student term papers can provide an account of the resources consulted in the research process, and a measure by which to evaluate undergraduate scholarship.

1.2. Verifying Citations: Accuracy and Persistency of URLs

In the world of academic scholarship, references form a link to original works, give credit to original ideas, and form a network of connections to related documents. They are also a means for checking against plagiarism. Unlike citations that refer to traditional paper books and journals, citations to Internet documents (URLs) are generally considered highly unstable.

In a study of 344 random URLs first collected in December 1996, only one URL (0.3%) from the sample was determined to still work 14 months later. Although the website itself has more permanence than its individual documents, the researcher estimated that the half-life of a website is approximately 2.9 years (Koehler, 1999).

Jill Watts, an Associate Professor of History at California State University at San Marcos was quoted recently regarding the decision on the part of Oxford University Press to not publish the 78-page bibliography associated with their edition of *The Nothing That Is: A Natural History of Zero* by Robert Kaplan with the printed book: "If what's on the Internet doesn't last forever, then we've lost power. If the academic apparatus disappears, then we're moving back to an earlier period when authors weren't citing sources and could just say whatever they wanted" (Carvajal, 2000).

The impermanence of URLs is not only a threat to scholarship, but imposes an unacceptably large burden on maintaining library catalogs (Shafer et al., 1996). There has been much work over the last several years developing standards for identifying and tracking Internet resources including URLs (Uniform Resource Locators); URNs (Uniform Resource Names); the OCLC Persistent URL (PURL); the SICI (Serial Item and Contribution Identifier); and the Digital Object Identifier (DOI) (Lynch, 1997).

This article presents a bibliometric analysis of undergraduate research papers and a verification of cited URLs. By comparing bibliographies from 1996 and 1999, the authors hope to identify changes in citation behavior during a period of widespread World Wide Web adoption on college campuses. Specifically, we were interested in answering the following questions: How has the composition of citations changed over time? Are students citing more resources in 1999 than in 1996? Are students using popular resources in lieu of scholarly ones? Are URLs still viable years after an Internet resource has been cited? What are the implications for higher education, and what actions (if any) should professors and librarians take to modify the research behaviors of students?

2. Methods

Introduction to Microeconomics (Econ 101) is a large freshman class taught to over 300 Cornell University students per year. Econ 101 is composed of students from the College of Arts and Sciences, the College of Agriculture and Life Sciences, and the School of Industrial and Labor Relations. As a term project, students are assembled into groups of four or five and are assigned a research question. Each group is expected to describe the problem in economic terms, find empirical data related to the economic principle, and provide an analysis of the findings. The project is a major component of their semester's work, and teams are expected to present their findings at the end of the course. Term papers are collected and archived by the professor to prevent "cribbing" from previous years' assignments. Three libraries on campus provide workshops on how to find information for the assignment. An online resource pathfinder is also provided.

In January 2000, 67 term papers from 1996 and 69 term papers from 1999 were collected from the professor. Bibliographies were photocopied and numbered for analysis, and any personal information was removed to preserve student confidentiality. Although the authors wished to study the relationship between grades and citation behavior, grades were not available for this study.

2.1. Bibliometric Analysis of Undergraduate Papers

Citations used in the bibliographies were coded based on the type of reference: Book, Journal, Magazine, Newspaper, Web, and Other. There was one category for Unidentifiable citations.

For the purposes of this study, Journals were defined as scholarly periodicals that contain primary research or substantial policy analysis. Examples of journals included: *The Quarterly Journal of Economics*, *Industrial and Labor Relations Review*, and the *Brookings Papers on Economic Activity*. Magazines were defined as nonscholarly periodicals that report primarily news, industry information, and events. Examples of magazines included: *Business Week*, *Fortune*, and *Pulp and Paper*. Although one might argue whether a serial might be considered a journal or a magazine, it was more important to be consistent with the coding for the purpose of comparing 1996 and 1999 data.

In 1999, many newspapers were available in print, from the publisher's web page, and through third party online providers like Lexis/Nexis. Because students may not have stated how they accessed the information, they were all coded as "newspapers." The same consistency was used for magazines and journals. No attempt was made to infer the source of the citation.

Chi-square tests were performed to identify differences between types (or categories) of references cited in 1996 and 1999. Although the assumption for independence among cases was not met (each reference is tied to an individual bibliography), this analysis was used anyway to better understand the data and should not be taken as strictly rigorous. The Student's *t*-test was also used to test the difference in means between 1996 and 1999.

Correlation analysis was performed to see if there were any tendencies for certain kinds of citations to be associated with each other (i.e., books with journals, and websites with magazines). One bibliography from 1996 was discovered to be a statistical outlier (reporting 33 citations) and was removed from several analyses to avoid skewing the data.

2.2. Verifying the Accuracy and Persistence of Internet Citations

During April and May 2000, Internet citations from 1996 and 1999 student bibliographies were checked for their accuracy and persistence. A citation was defined as an Internet resource if a URL was included and/or if the reference indicated "WWW," "Internet," or "Online."

There were two initial categories set up for defining the persistence of Internet citations: (1) the URL leads directly to the cited document; and (2) the URL does not lead directly to the cited document. The second category was further divided into three subcategories: (a) the document was found at a different URL; (2) the URL cited contains a typo; or (3) the document was not found at all.

If the URL did not correctly point to the cited document, attempts were made to determine if the document was still accessible on the Internet. URLs were first checked for obvious typographical errors. If no typographical errors were detected, then the URL was typed in, removing one directory level at a time, until a working web page was found. This page was examined for any link to the cited document. If the cited document was still not found, the home page for the site was located and various techniques (site maps, internal search engines, etc.) were used to locate the document on the server.

If this strategy did not work, then an Internet search engine, *Google*, was used to try to locate the document. If Google did not return the document on the first screen of results, then the document was considered to be inaccessible on the Internet. If there was no title or author given in the bibliographic reference (only a URL), it was impossible to search for the document, with a few exceptions.

3. Results

3.1. Bibliometric Anaysis

It was thought that the number of references cited per bibliography would increase as access to electronic resources became more available in the late 1990s. Surprisingly, the mean number of citations remained virtually unchanged (11.3 in 1996 to 11.9 in 1999). The mean number of books, newspapers, and websites in bibliographies did change significantly ($P < 0.001$) from 1996 to 1999. There were no significant changes with respect to journals or magazines (Table 1). The median number of citations, a better indication when the data are skewed, increased from 10 in 1996 to 12 in 1999.

In general, students cited fewer books in 1999 than they did in 1996. Comprising nearly one third (30%) of total citations in 1996, book citations dropped to less than one fifth (19%) in 1999. This translated into a decrease from 3.5 books per bibliography in 1996 to 2.2 in 1999, with the median citation number dropping from 3 to 1. Not surprisingly, those who cited only one book cited the course textbook. In 1996, only 17% of bibliographies included no citations to books compared to 29% in 1999. Nineteen percent of 1996 bibliographies contained more than six book citations, compared to only 3% in 1999.

Although websites became a dominant category in 1999 comprising 21% of all citations, there was relatively little change in the composition of cited domains. From 1996 to 1999, percentages of cited domains changed as follows: .com (39% to 36%), .edu (7% to 8%), .org (29% to 35%), .gov (25% to 21%), and .net (0% to 1%).

3.1.1. A Diversion from Scholarly Resources

To confirm a speculation that students were moving from scholarly to nonscholarly resources, book citations were combined with journal citations to form a category called "Scholarly." Newspapers and magazines were also combined to form a category called "Nonscholarly." Because

TABLE 1. Descriptive statistics by genera.

	Total citations	Books	Journals	Magazines	Newspaper	Web	Other	Unidentifiable
1996								
Mean	11.3	3.5*	2.6	2.7	0.9*	1.1*	0.35	0.11
Median	10	3	2	2	0	0	0	0
SD	4.03	2.79	2.60	2.76	1.81	1.92	1.12	0.53
Min	4	0	0	0	0	0	0	0
Max	22	11	9	10	9	9	8	4
Sum	733	230	171	176	57	69	23	7
%	100	30	22	27	7	9	3	1
1999								
Mean	11.9	2.2*	2.4	2.5	1.9*	2.5*	0.2	0.12
Median	12	1	1	1	1	1	0	0
SD	4.45	2.29	3.12	3.03	2.71	2.84	0.58	0.40
Min	2	0	0	0	0	0	0	0
Max	25	9	16	15	12	14	3	2
Sum	818	152	166	174	130	175	14	8
%	100	19	20	21	16	21	2	1

Asterisk denote a significant difference in means between 1996 and 1999.

many of the web citations do not presently work, and those that do work are mutable and difficult to judge, web documents were not used as part of this measure. The difference in "Scholarly" and "Nonscholarly" means was tested using the Student's *t*-test, where each "case" was a bibliography.

There was a significant decline in the mean use of scholarly materials from 6.1 citations per bibliography in 1996 to 4.6 in 1999 ($P < .02$). This change can be explained almost entirely by the decrease in book citations (see Table 1). There was no corresponding significant increase in the mean use of nonscholarly materials; however, it should be noted that web citations were not factored into the analysis. If they had been, we might have seen a significant change.

Correlation analysis was performed to see if there were any tendencies for certain kinds of citations to be associated with each other (i.e., books with journals, and websites with magazines). No strong correlations were found.

3.2. Accuracy and Persistency of URLs

In general, there was poor success with retrieving web documents that were cited in student bibliographies. For

1999 bibliographies, 55% (109) of URLs lead directly to the cited document. Nineteen percent (37) of the cited documents were found at a different URL, and 10% (19) of citations contained typographical errors. Sixteen percent (32) were not found at all. For 1996 bibliographies, only 18% (13) of URLs lead directly to the cited document, 26% (19) of the cited documents were found at a different URL, and 3% (2) of citations contained typos. Fifty-three percent of cited documents were not found at all (Figs. 1 and 2).

4. Discussion

Although this study of one undergraduate class at Cornell University cannot be considered representative of college students in general, there has been a remarkable shift in the use of information resources away from books in favor of general web resources. Universities with large library collections—often a measure by which research universities are compared—should be concerned if students are no longer taking the opportunity to use them. Professors should
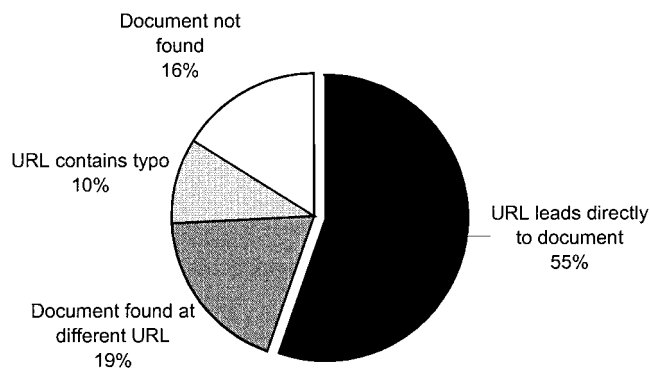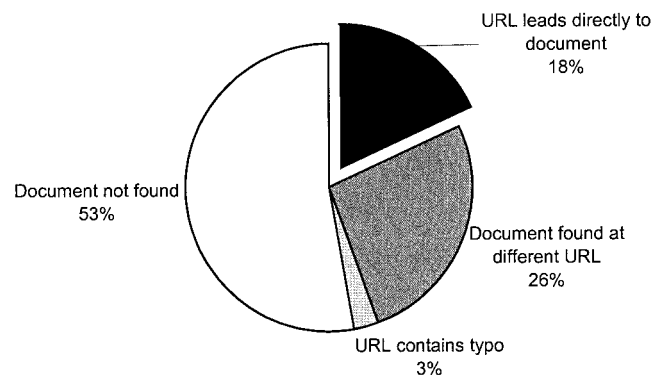

FIG. 1. Persistency of URLs cited in 1999 ($N = 197$).

FIG. 2. Persistency of URLs cited in 1996 ($N = 72$).

be concerned that they are not exposing their students to academic literature in their field.

The wide availability of electronic resources (both subscription-based services and free websites) has facilitated the access of some materials that have been traditionally difficult to locate. The rise in newspaper citations in 1999 may be explained by two contributing factors. In 1997, the Cornell University Library began a campus-wide subscription to Lexis/Nexis Academic Universe, an online resource that provides full-text access to hundreds of newspapers. Many major newspapers have also been offering top stories free from their websites.

Since 1996, the number of online magazines and journals has also increased greatly. Products that aggregate and provide full-text access to hundreds of journals (like ProQuest Direct) have been available for several years. In the last two years, academic publishers have been providing electronic journal subscriptions to libraries.

Along with electronic access of materials comes the issue of convenience. In the current information environment of electronic and print, electronic has a pronounced advantage with regards to ease of access. This might be especially true for students, many of whom work on their papers the night before they are due.

Academic libraries have been actively acquiring scholarly materials and providing campus-wide access via library gateways. In the case of the Cornell University Library, the Library Gateway (http://campusgw.library.cornell.edu) offers access to hundreds of online periodicals, indexes, numeric data, and other scholarly electronic resources. Currently, the library is testing access to full-text books.

Although the *ideal* fully-networked electronic academic library would have a positive effect on the information use of our students, we need to realize that the library does not have complete control over the information available to students, and competes with other resources available on the Internet as a whole. Providing an alternative to surfing the Internet is only part of the solution.

4.1. What Should Be Done?

Professors and librarians alike know that students are very literal with the requirements of their term-paper assignment. Journals that are merely suggested as starting points for assignments are often considered by students to be requirements of the research. The authors argue that professors ought to be more prescriptive with the type of literature that they would like to see consulted. Many professors already stipulate that a minimal number of peer-reviewed journal articles or books need to be used for an assignment. Some professors state that they do not accept websites as bona fide references, but the prohibition of web citations might be considered arcane, considering the trend in academic web publishing.

Lubans argues that a single electronic portal for scholarly research is required for students. Libraries have been experimenting with including scholarly websites in their on-line catalogs. In 1999, OCLC began developing a union-type catalog called the Cooperative Online Resource Catalog (CORC). CORC was designed to share metadata, and with it, the responsibility of cataloging and sharing Internet resources and pathfinders. To address the impermanence of URLs, CORC incorporates a maintenance feature that checks for broken or redirected links. OCLC hopes that this system will replicate the success of WorldCat (a shared bibliographic database), and enable libraries to develop their own gateways to online resources (Chepesiuk, 1999).

McBride and Dickstein argue from a more pedagogical approach. They maintain that "the first step for academics is to teach students how to find information from all scholarly sources, whether print or on line. The second step is to teach students how to read that material critically, even suspiciously" (McBride & Dickstein, 1998). A renewed push for teaching the evaluation of sources began as early as 1995—about the same time that the Web was gaining widespread popularity on college campuses. Librarians quickly developed evaluation checklists and tutorials for their students, with many of them on the Web (Auer, 1998).

Increased guidance by librarians is not the only solution to the declining use of scholarly resources. The authors argue for a multipart solution that includes (1) setting stricter guidelines for acceptable citations in course assignments, (2) creating and maintaining scholarly portals for authoritative websites with a commitment to long-term access, and (3) continuing to instruct students how to critically evaluate resources. Taken together, this solution provides the authority, responsibility, and accountability for both professors and librarians in the education of college students.

References

ACRL (2000). Association of research libraries statistics [On-line]. University of Virginia Library. Available: http://fisher.lib.virginia.edu/newalr/.

Auer, N. (1998). Bibliography on evaluating Internet resources. Teacher Librarian, 25, 23–24.

Available: http://purl.oclc.org/OCLC/PURL/INET96

Carvajal, D. (2000). The book's in print, but its bibliography lives in cyberspace. NY Times, May 29, Late Edition, A1.

Chepesiuk, R. (1999). Organizing the Internet: The "core" of the challenge. American Libraries, 30, 60–63.

D'Esposito, J.E., & Gardner, R.M. (1999). University students' perceptions of the Internet: An exploratory study. Journal of Academic Librarianship, 25, 456–461.

Koehler, W. (1999). An analysis of web page and web site constancy and permanence. Journal of the American Society for Information Science, 50, 162–180.

Lubans, J. (1998). How first-year university students use and regard Internet resources [On-line]. Available: http://www.lib.duke.edu/lubans/docs/1styear/firstyear.html

Lubans, J. (1999). When students hit the surf. School Library Journal, 45, 144–147.

Lynch, C. (1997). Identifiers and their role in networked information applications. Bulletin of the American Society for Information Science, 24, 17–20.

Majewski, K. (1999). The Internet and the undergraduate research paper. Unpublished report, SUNY College of Technology at Delhi.

Malone, D., & Videon, C. (1997). Assessing undergraduate use of electronic resources: A quantitative analysis of works cited. Research Strategies, 15, 151–158.

McBride, K.B., & Dickstein, R. (1998). The Web demands critical reading by students. The chronicle of higher education, 44, B6.

OCLC CORC: Cooperative online resource catalog [On-line]. Available: http://www.oclc.org/oclc/corc/

Rothenberg, D. (1998). How the Web destroys student research papers. The Education Digest, 63, 59–61.

Shafer, K., Weibel, S., Jul, E., & Fausey, J. (1996). Introduction to persistent uniform resource locators [On-line]. OCLC. Available: http://purl.oclc.org/OCLC/PURL/INET96