Effect of the Web on Undergraduate Citation Behavior: Guiding Student Scholarship in a Networked Age

Philip M. Davis

abstract: This article provides the last update to a longitudinal study tracking the research behavior of a multi-college undergraduate course in microeconomics from 1996 to 2001. Student term paper bibliographies grew between 1996 and 2000 but included fewer scholarly resources. In 2001, students tended to cite scholarly sources when the professor provided clear and enforceable guidelines in his class assignment. The accuracy and persistency of cited Web documents also increased as a result.

T he wiring of the U.S. college campus has had a pronounced 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 and dorm rooms. As a consequence, libraries have been experiencing a downward (and accelerating) trend in the number of questions asked at reference desks. According to annual statistics compiled by the Association of Research Libraries, the median number of reference questions asked in libraries has dropped by 30 percent between 1995 and 2001. Cornell University, the author's institution, has experienced a 44 percent decline.[1]

A recent survey of 3,200 students and faculty at universities and liberal arts colleges was sponsored by the Digital Library Federation and conducted by Outsell, Inc. Preliminary results of the survey reported that of the time students devote to searching for information, roughly one-third is spent in the campus libraries, versus one-half from their residences.[2]

During the period of Internet adoption on college campuses and the subsequent development of the graphical Web browser, student research papers have reflected their preference for networked information. In their first article tracking a large multi-college class in microeconomics from 1996, the author and Suzanne Cohen noted a dra-

matic shift from the use of credible, peer-reviewed materials to popular and unfiltered information in term paper bibliographies.[3] The following year, even after the professor provided verbal guidelines on using appropriate scholarly materials, student bibliographies continued to get larger and contain more unscholarly materials.[4]

During the period of Internet adoption on college campuses and the subsequent development of the graphical Web browser, student research papers have reflected their preference for networked information.

A preliminary study of student papers from six subject disciplines at the College of Mount St. Joseph indicated that some fields of study may be more prone to student use of Web sources, specifically, the Humanities and Social Sciences.[5] The study was too small to determine whether the difference was dependent upon the individual instructor (as some had explicit guidelines for use of Web sources) or the class assignment. Anecdotal evidence from professors and librarians suggest 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.

Deborah Grimes and Carl Boening evaluated the kinds of resources students are citing in introductory English composition classes, and interviewed both students and faculty for their perceptions on citing Internet resources. Not surprisingly, they found that students are using unevaluated resources, and a gap exists between what professors expect and what students actually use. The authors concluded that students were either ill-equipped or unwilling to make the effort to evaluate Web resources.[6]

In an exploratory focus group of undergraduate perceptions of the Internet, Joann D'Esposito and Rachel Gardner report that students were keenly aware of the importance of discerning the difference between reliable and unreliable information from the Internet. Students reported that the Internet sites of highest quality and reliability were those produced by the government, educational institutions, and reputable businesses and corporations.[7]

Convenience may play a large role in the selection of Web resources over traditional print. In the current mixed environment of electronic and print literature, electronic literature has a pronounced advantage because it is easy to access. This might be especially true for students, many of whom work on their papers the night before they are due. A survey of 543 college students at Iowa State University reported that 63 percent of respondents ranked most highly those resources that were easy to use and to find, whether those sources were available from the library or the Internet.[8] At Duke University, a survey of 1,200 freshman indicated that more than half of the students believed that the Web is a time-saver when they search for information.[9]

How do professors feel about students using the Web for their research? Susan Davis Herring surveyed faculty acceptance of the Web for student research and concluded that faculty generally feel positive about using the Web as a research tool; however, they question the accuracy and reliability of Web content. Faculty are chiefly concerned about students' ability to evaluate the information they find on the Internet, and

their concerns are not without justification. [10] An analysis of Internet sources cited by students in a literature class discovered that 42 percent of their citations pointed to other student papers on the Web.[11]

Over the last few years, many articles have suggested how both professors and librarians can assist students in evaluating properly what they find on the Web.[12] Kari Boyd McBride and Ruth Dickstein argue from a 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 online. The second step is to teach students how to read that material critically, even suspiciously."[13]

This article builds upon the above research by summarizing the Cornell study tracking undergraduate citations from 1996 to 2001. It analyzes the effects professors and librarians have on student research and provides suggestions for best practices.

Methodology

Introduction to Microeconomics (Econ 101) is a large class taught to more than 1,200 Cornell University students per year. The Fall lecture that provides the source data for this paper normally enrolls about 350 students. Econ 101 is composed primarily of students from Cornell's College of Arts and Sciences, the College of Agriculture and Life Sciences, the College of Human Ecology, and the School of Industrial and Labor Relations. As a term project, students are randomly assembled into groups of four or five and are assigned a random, instructor-written research question based on the contemporary business and economic press. 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. The source topics are not re-used and are based on a newspaper or magazine article published within the last year. Three libraries on campus provide workshops on how to find information for the assignment. An online resource pathfinder (bibliography) is also provided.

Citation Guidelines

As part of the assignment, the professor provided guidelines on acceptable research sources. The following statement was given from 1996 onward, with 2001 additions indicated with italics.

> You must provide a complete, correct citation for any material you use in preparing your report. *At least five of your sources must be published, scientific (peer-reviewed or university press) articles or pre-prints. Even if you use electronic sources for these articles, you must provide a proper bibliographic citation.* A complete citation consists of a standard bibliographic entry (for printed material) or an equivalent, certified location for computer stored data items. A uniform resource locator (URL) is not sufficient. You may use confidential or unpublished material in your report if, and only if, you provide us with a copy of this information and the name, address, and telephone number of your source. *TAs have been*

instructed to check all citations and will penalize projects with inadequate scientific references or incomplete bibliographic citations.

The additions in 2001 provided explicit parameters for the students. A minimum number of citations was specified, along with a standard for electronic and nontraditional resources, followed by a warning with consequences for those students who don't follow the guidelines.

Citation Analysis

Term papers were submitted to the professor in print format for years 1996 and 1999. Beginning in 2000, students submitted papers in electronic format. There were sixty-seven papers submitted in 1996, sixty-nine in 1999, sixty-three in 2000, and sixty-nine in 2001. Bibliographies from these papers were stripped of personal information to preserve student confidentiality.

Citations 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 non-scholarly periodicals that report primarily news, industry information, and events. Examples of magazines included: *Business Week*, *Fortune*, and *Pulp and Paper*. While 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 making yearly comparisons.

By the late 1990s, many journals, magazines, and newspapers were available in print, from the publisher's Web page, and through third party online providers like Lexis/Nexis. Since students may not have stated how they accessed the information, all traditional print materials were coded as such even if they may have been accessed electronically. No attempt was made to infer the source of a citation. Web resources were identified as electronic-only resources with no print counterpart.

Chi^2 tests were performed to identify differences among types (or categories) of references cited in 1996, 1999, 2000, and 2001. Analysis of Variance (ANOVA) was also used to test the difference in means between years.

Verifying the Accuracy and Persistence of Internet Citations

Internet citations from the bibliographies were checked for accuracy and persistence six months after the students submitted the papers.[14] A citation was defined as an Internet resource if a URL was included and/or if the reference indicated "WWW," "Internet," or "Online."

We set up two initial categories 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 sub-categories: a) the document was found at a different URL; b) the URL cited contains a typo; or c) the document was not found at all.

If the URL did not correctly point to the cited document, we attempted to determine if the document was still accessible on the Internet. We checked URLs for obvious typographical errors. If no typographical errors were detected, we typed in the URL, removing one directory level at a time, until a working Web page loaded. We examined this page for any link to the cited document. If the cited document still did not emerge, we searched for the home page for the site and used various techniques (site maps, internal search engines, etc.) to locate the document on the server.

If this strategy did not work, we used an Internet search engine, Google, to try to locate the document. If Google did not return the document on the first screen of results, 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), searching for the document was impossible, with a few exceptions.

Observations

The following section describes what students cited in their term papers, how these citations were distributed between types of materials, the domain of cited Web documents, and the persistency of these Web documents over time.

Composition of Citations

The number of book citations in student bibliographies dropped significantly (P<0.001) from 1996 to 2001, from 30 percent of cited sources to 16 percent. Journal citations remained relatively constant for the first three years tracked and then rose dramatically in 2001 when the professor set minimum requirements for scholarly sources. Magazine use dropped slightly (nonsignificantly) from 1996 and then continued to remain relatively constant from 1999. Newspaper use increased significantly after 1996.[15] The "Other" category remained relatively constant over the years, with the exception of 2001. In 2001, this group was mostly composed of government reports (including Hearings, Acts of Congress, and the Code of Federal Regulations). The Web category initially showed a significant jump from 1996 to 1999 as wide access to the Internet was established in student dorms, but fell almost as dramatically in 2001 as the professor provided guidelines on appropriate research sources.

> **The Web category initially showed a significant jump from 1996 to 1999 as wide access to the Internet was established in student dorms, but fell almost as dramatically in 2001 as the professor provided guidelines on appropriate research sources.**

Distribution of Citations

Looking at the average of all references cited in bibliographies is useful; however, there is no "average" bibliography. By looking at the distribution of term paper citations
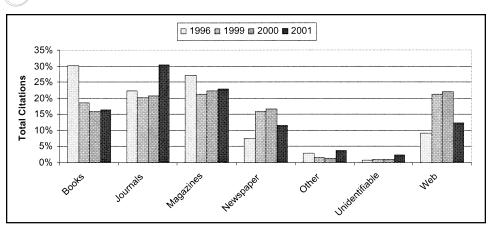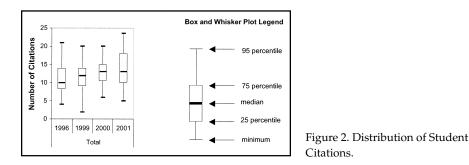
Figure 1. Composition of Citations in Student Bibliographies 1996, 1999, 2000, and 2001.

(Figures 2 and 3), we can understand better the citation behavior of the class. Box and Whisker Plots provide us with a visual description of the data. On each end (the whisker) the minimum and maximum data points are plotted. In our graph, the 95th percentile was used instead of maximum since the data often included outlier data. The "box" indicates the 25th and 75th percentile, with the median (50th percentile) represented as a horizontal bar. The box is the most informative feature since it represents 50 percent of all citations, and gives an indication of the skew of the distribution.

Bibliographies Getting Bigger

Between 1996 and 2001, there was a significant increase in the number of citations per bibliography (P<0.01). The average number of citations per bibliography changed from 11.3 in 1996 to 14.4 in 2001. The median number of citations per bibliography increased from ten in 1996 to thirteen in 2000 and 2001 (figure 2). There was, however, a greater range in the number of citations used in 2001. The 75th percentile for 2001 was eighteen citations/bibliography. In other words, 25 percent of the student papers included more than eighteen citations.

Figure 2. Distribution of Student Citations.

Returning to Scholarly Bibliographies

In 2001, after faculty became concerned that student papers were becoming less scholarly (i.e. including fewer books, and including more unclassified Web sites and news sources), the professor implemented written and enforceable guidelines for acceptable reference sources. As a result, the 2001 bibliographies were remarkably different from previous years. Book citations rose, journal citations increased dramatically, and Web citations decreased along with newspaper citations (figure 3). The number of scholarly citations (the culmination of books and journals) returned to 1996 levels. What is apparent from this graph is that while students are citing as many scholarly materials as they did in 1996, they are now supplementing these sources with nonscholarly materials from the Web.
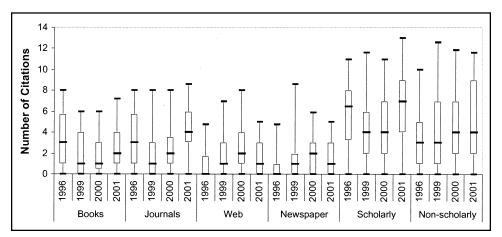
> **What is apparent from this graph is that while students are citing as many scholarly materials as they did in 1996, they are now supplementing these sources with nonscholarly materials from the Web.**

Figure 3. Distribution of Student Citations by Genre.

Web Citations

Web citations comprised 9 percent of citations in 1996, swelled to 21 percent and 22 percent in 1999 and 2000, and then declined to 13 percent in 2001 (figure 1). Since 1996, there has been relatively little change in the composition of cited domains (Figure 4). The dot-coms continue to be the mostly heavily cited category.
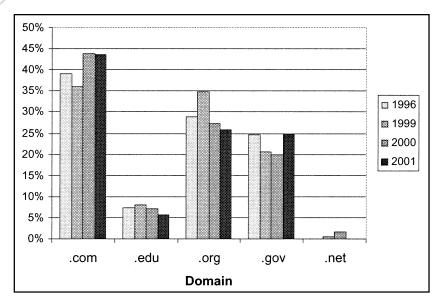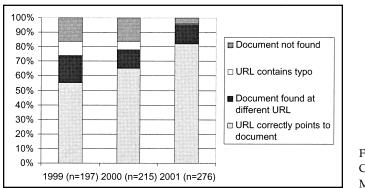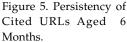
Figure 4. Composition of Cited URLs.

Persistency of URLs

When URLs cited in 1996 were first checked for their persistency in 1999 (3.5 years after they were cited), only 18 percent of them pointed directly to the cited Internet document. Twenty-six percent were found at a different URL, 3 percent contained typos, and more than half (53 percent) could not be found at all after performing multiple searches and browses.[16]

The accuracy of cited URLs has increased over the last three years (figure 5). In 1999 55 percent of the citations correctly pointed to the document on the Internet after being aged for six months. This increased to 65 percent in 2000 when the professor began requiring students to submit their papers electronically. For 2001 citations, 82 percent correctly point to the document. Internet documents continue to change locations (between 13 and 19 percent), even within six months. In 2001, only one of the URLs cited (0.4 percent) contained an apparent typographical error.

Figure 5. Persistency of Cited URLs Aged 6 Months.

Discussion

This article tracks the citation behavior of a single multi-college class at an Ivy League university. While acknowledging its limitations, the author believes that the results of this study are generalizable to other institutions. The professor's solution to the problem in undergraduate research was not to ban Web-based citations, but to provide acceptable parameters for their use. The results of this study indicate that students will meet the expectations of the professor when those expectations are clearly articulated and enforced.

Why is it important that professors provide research parameters in their assignments?

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. Because of this loss of monopoly, professors can no longer solely rely on the library to serve as the intermediary between the student and the universe of information. The networked environment shifts much of this responsibility to the professor.

Setting minimum guidelines in assignments ensures that students will attempt to identify relevant scholarly literature in their subject field. It helps students develop the skills necessary to distinguish scholarly resources

> **From the perspective of the library, minimum scholarship guidelines affirm the value of a library's collection and lend more relevance to library-mediated instruction.**

from popular ones and gives students the ability to choose from a multitude of sources without the professor being unduly prescriptive. From the perspective of the library, minimum scholarship guidelines affirm the value of a library's collection and lend more relevance to library-mediated instruction.

Why are bibliographies getting bigger?

Access to information is not a limiting factor to student research—time is. Students, many of whom are working on their term papers the night before they are due, may be selecting Internet resources because they perceive them to be more convenient than traditional library research. From 1996 to 1999, we observed a significant increase in the size of bibliographies. The increase was explained entirely by an increase in traditionally nonscholarly citations (newspapers and Web sites), in spite of the fact that scholarly citations (books and journals) decreased during this same period. When the professor implemented minimum standards in his term paper assignment, the size of bibliographies did not change substantially from the previous year, although the number of scholarly citations did return to 1996 levels.

The change in citation composition may also reflect the change in accessibility of sources that have been traditionally difficult for student research: news, company, and government information. These sources are now available from library-purchased sub-

scription databases or available directly from news, business, and government Web sites.

Scholarly research has never been (and probably will never be) as simple as doing an Internet search. The increase in size of student bibliographies may simply reflect the ease and speed of Internet search engines, and explain why students appear to be supplementing their scholarly resources with nontraditional resources from the Web.

Why did Web citations get more accurate, and why is accuracy important in scholarly research?

The accuracy of cited URLs aged six months has steadily increased since 1999. Part of this increase in accuracy may be because the professor required students to submit their papers electronically beginning in 2000, which resulted in fewer typographical errors. In 2001, the professor made it explicit that students would be penalized for incorrect or incomplete citations.

An additional explanation is that students in this class have become more selective of what they cite in their papers and are selecting documents that are more inherently stable. The long-term persistency of a cited URL is of great concern to researchers, since the citation forms the basis of academic scholarship.

Vicki Burton, a professor of English, and Scott Chadwick, a professor of Journalism and Communication, write, "Whether student researchers are choosing inappropriate sources due to lack of training, lack of time, lack of discretion, or for some other reason, the practice merits attention because it both devalues and places at risk a central assumption of academic writing: that a writer will support claims with appropriate, valid and authoritative evidence."[17]

This evidence comes down to the longterm persistence of the citation. In the world of scholarship, references form a link to original works, give credit to original ideas, and form a network of connections to related documents. A viable link—whether in print or electronic form—is absolutely necessary in order to preserve scholarly communication. Without citations that pass the test of time, we have no way to proceed forward because we can no longer see the past.

Acknowledgements

Philip M. Davis is the Life Sciences Bibliographer at Cornell University; he may be contacted via e-mail at: pmd8@cornell.edu.

Notes

1. Association of Research Libraries, "ARL Statistics Interactive Edition" (University of Virginia Library, 2002) cited April 2, 2002. Available: <http://fisher.lib.virginia.edu/arl/index.html> [October 10, 2002].

2. Daniel Greenstein and Leigh Watson Healy, "National Survey Documents Effects of Internet Use on Libraries," *CLIR Issues* 27 (2002). Available: <http://www.clir.org/pubs/issues/issues27.html> [October 10, 2002].

3. Philip M. Davis and Suzanne A. Cohen, "The Effect of the Web on Undergraduate Citation Behavior 1996–1999," *Journal of the American Society for Information Science and Technology* 52, no. 4 (2001): 309–14.

4. Philip M. Davis, "The Effect of the Web on Undergraduate Citation Behavior: A 2000 Update," *College & Research Libraries* 63, no. 1 (2002): 53–60.

5. Paul O. Jenkins, "They're Not Just Using Web Sites: A Citation Study of 116 Student Papers," *College and Research Libraries News* 63, no. 3 (2002): 164.

6. Deborah J. Grimes and Carl H. Boening, "Worries with the Web: A Look at Student Use of Web Resources," *College & Research Libraries* 62, no. 1 (2001): 11–23.

7. Joann E. D'Esposito and Rachel M. Gardner, "University Students' Perceptions of the Internet: An Exploratory Study," *Journal of Academic Librarianship* 25, no. 6 (1999): 456–61.

8. Vicki Tolar Burton and Scott A. Chadwick, "Investigating the Practices of Student Researchers: Patterns of Use and Criteria for Use of Internet and Library Sources," *Computers and Composition* 17, no. 3 (2000): 309–28.

9. John Lubans, *How First-Year University Students Use and Regard Internet Resources* (1998). Available: <http://web.archive.org/web/20011105120257/http://www.lubans.org/docs/1styear/firstyear.html >[October10, 2002].

10. Susan Davis Herring, "Faculty Acceptance of the World Wide Web for Student Research," *College & Research Libraries* 62, no. 3 (2001): 251–58.

11. Mary Ann Gillette and Carol Videon, "Seeking Quality on the Internet: A Case Study of Composition Students' Works Cited," *Teaching English in the Two Year College* 26, no. 2 (1998): 189–94.

12. Catherine B. Elliott, "Helping Students Weave Their Way through the World Wide Web," *English Journal* 90, no. 2 (2000): 87–92; Susan A. Gardner, Hiltraut H. Benham, and Bridget M. Newell, "Oh, What a Tangled Web We've Woven! Helping Students Evaluate Sources," *English Journal* 89, no. 1 (1999): 39–44; Manfred Kuechler, "Using the Web in the Classroom," *Social Science Computer Review* 17, no. 2 (1999): 144–61; Delia Neuman, "What Do We Do after the School Has Been Wired? Providing Intellectual Access to Digital Resources," *NASSP Bulletin* 83, no. 605 (1999): 35–43; Robert N. Ronau, Steve Ryan, and Beth Stroble, "Searching the Web: New Tools—New Dilemmas," *Internet and Higher Education* 1, no. 4 (1998): 253–62; Jean Shackelford, Dot S. Thompson, and Mary Beth James, "Teaching Strategy and Assignment Design: Assessing the Quality and Validity of Information Via the Web," *Social Science Computer Review* 17, no. 2 (1999): 196–208; Madeleine Sorapure, Pamela Inglesby, and George Yatchisin, "Web Literacy: Challenges and Opportunities for Research in a New Medium," *Computers and Composition* 15, no. 3 (1998): 409–24; Nicole Auer, "Bibliography on Evaluating Internet Resources," *Teacher Librarian* 25, no. 5 (1998): 23–24; Ross J. Todd, "Transformational Leadership and Transformational Learning: Information Literacy and the World Wide Web," *NASSP Bulletin* 83, no. 605 (1999): 4–12.

13. Kari Boyd McBride and Ruth Dickstein, "The Web Demands Critical Reading by Students," *The Chronicle of Higher Education* 44, no. 28 (1998): B6.

14. In the first study by Davis and Cohen, citations from 1996 papers were checked three and a half years after they were submitted.

15. The most probable explanation for this increase was the introduction of a campus-wide subscription to Lexis/Nexis, a full-text news resource, in 1997.

16. Davis and Cohen, "The Effect of the Web on Undergraduate Citation Behavior 1996–1999," 309–14.

17. Burton and Chadwick, 310.