

# The Effect of the Web on Undergraduate Citation Behavior: A 2000 Update

Philip M. Davis

This paper provides a 2000 update to the 1996–1999 citation analysis of undergraduate term papers by Philip M. Davis and Suzanne A. Cohen.<sup>1</sup> The total number of bibliographic citations continued to grow from a median of ten in 1996 to thirteen in 2000. However, this growth is entirely explained by the addition of traditionally nonscholarly materials (Web and newspaper citations). A significant improvement in the accuracy of Internet citations was found when term papers were submitted electronically. In 2000, the first year of electronic submissions, 65 percent of the citations pointed directly to the cited document, up from 55 percent in 1999. Internet citations aged six months in both 1999 and 2000 bibliographies were still irretrievable anywhere on the Internet 16 percent of the time. If more scholarly citations in term papers are to be seen, professors must provide clear expectations in their class assignments. Students should be required to submit an electronic copy of their paper so that Internet citations can be scrutinized for accuracy and plagiarism.

**I**n the first analysis of undergraduate term paper bibliographies, Davis and Cohen documented a significant decrease in the frequency of traditional scholarly resources cited between 1996 and 1999.<sup>2</sup> The authors recommended (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 on how to critically evaluate resources.

After the results were known, the authors consulted with the professor who was determined to change the following year's class. The professor met with his

teaching assistants and instructed them that he wanted to see more scholarly resources used in term papers. He also instilled the same need with librarians organizing the library research sessions with the students. In 2000, students were required to submit their term papers electronically. The only element that did not change was the wording in the term paper assignment, which stayed exactly the same from 1999.

This article compares the citations found in the 2000 term paper bibliographies with those submitted in 1996 and 1999. It tests the assumption that papers submitted electronically would exhibit higher accuracy when citing Internet citations. It then discusses whether the

---

*Philip M. Davis is the Life Sciences Bibliographer in the Albert R. Mann Library at Cornell University; e-mail: pmd8@cornell.edu. The author wishes to thank Suzanne Cohen for her invaluable help as a reviewer and collaborator in this ongoing study.*

professor's verbal guidelines on scholarly research had an effect on the contents of student bibliographies.

### Literature Update

Over the past few years, considerable anecdotal evidence has suggested that students prefer electronic resources, lack the ability or willingness to distinguish credible academic sources from popular materials on the Internet, and have difficulty citing what they find. Articles confirming these observations have recently appeared in print.

In an exploratory focus group of undergraduate perceptions of the Internet, Joann E. D'Esposito and Rachel M. Gardner reported that students were keenly aware of the importance of discerning reliable 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.<sup>3</sup>

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, yet question the accuracy and reliability of its content. Faculty are chiefly concerned about students' ability to evaluate the information they find on the Internet.<sup>4</sup>

Deborah J. Grimes and Carl H. 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.<sup>5</sup> Not surprisingly, they found that students are using unevaluated resources and that 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.

As solutions, these articles mention library instruction, bibliographies (print or online), and class support materials, yet none mention the need to include guidelines, examples, or minimum research

standards *in* class assignments. Unfortunately, students often discount library instruction when it is offered as an add-on to courses because they view it as extrinsic to their course work and irrelevant to their grades.<sup>6</sup>

Having input on course research assignments is how librarians can make the most difference. In 1998, the ACRL Task Force on Academic Library Outcomes Assessment created a report that developed principles, standards, and recommendations for outcome-based evaluation of library instruction.<sup>7</sup> The report recommended that "syllabi and course assignments [should] include information [for] literacy skills development" and advised using "course assignments and syllabi analysis" as an evaluation method to determine success.

### Methodology

Introduction to Microeconomics (Econ 101) is a large freshman class taught to more than three hundred Cornell University students each 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 (bibliography) also is provided.

Sixty-seven term papers from 1996 and sixty-nine papers from 1999 were collected from the professor. Sixty-three papers submitted electronically for the 2000 class were sent to the researchers digitally.

Bibliographies were stripped of personal information to preserve student confidentiality. Grades for the 2000 papers were available for analysis.

**Bibliometric Analysis of Undergraduate Papers**

Citations used in the bibliographies were coded based on type of reference: book, journal, magazine, newspaper, Web, and other. One category was 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 whether a serial might be considered a journal or a magazine is arguable, 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 such as Lexis/Nexis. Because students may not have stated how they accessed

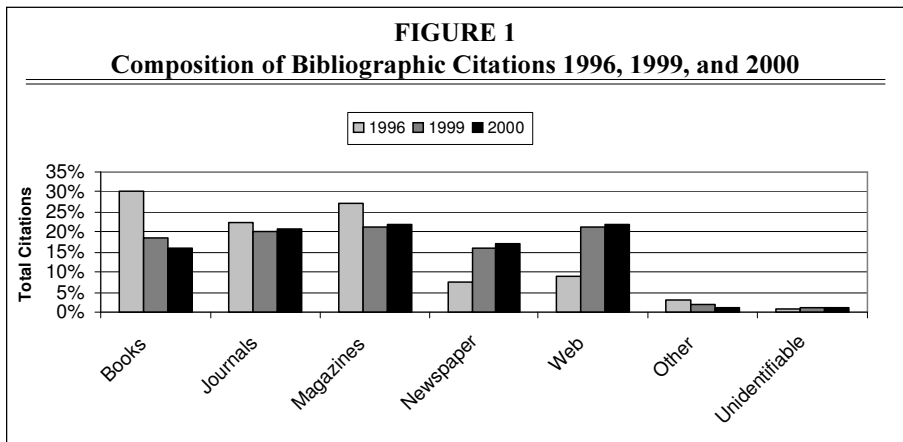
the information, all traditional print materials were coded as such even if they might have been accessed electronically. No attempt was made to infer the source of a citation. Web resources were identified as electronic-only resources with no print counterpart.

**Looking at the average of all references cited in bibliographies is useful, but, in reality, there is no “average” bibliography.**

Chi-square tests were performed to identify differences among types (or categories) of references cited in 1996, 1999, and 2000. Although the assumption for independence among cases was not met (each reference is tied to an individual bibliography), this analysis was used anyhow to better understand the data and should not be taken as strictly rigorous. Analysis of variance (ANOVA) also was used to test the difference in means between 1996, 1999, and 2000. Finally, regression analysis was used to see whether there was a relationship between citation behavior and grade.

**Verifying the Accuracy and Persistence of Internet Citations**

Internet citations from the 2000 bibliographies were checked for accuracy and persistence six months after the papers were submitted. A “citation” was defined



as an Internet resource if a URL was included and/or if the reference indicated WWW, Internet, or Online.

Two initial categories were set up for defining the persistence of Internet citations: the URL leads directly to the cited document; and the URL does not lead directly to the cited document. The second category was further divided into three

subcategories: the document was found at a different URL; the URL cited contains a typo; or the document was not found at all.

If the URL did not correctly point to the cited document, attempts were made to determine whether the document was still accessible on the Internet. URLs were first checked for obvious typographical errors. If no typographical errors were detected, 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, 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, the document was considered to be inaccessible on the Internet. If no title or author was given in the bibliographic reference (only a URL), it was impossible to search for the document, with a few exceptions. Focusing on professional publications in computer science, Steve Lawrence et al. employed a similar method of searching and browsing for incorrect URLs.<sup>8</sup>

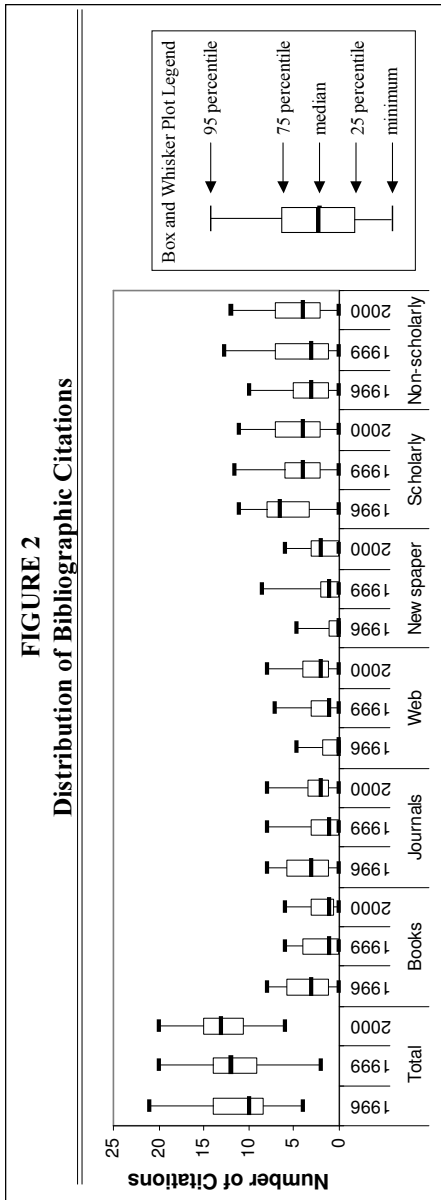
**Results**

*Composition of Citations*

The composition of citation categories from 1999 to 2000 remained virtually the same, with only slight (statistically insignificant) decreases in books and corresponding increases in other categories. There still remains a significant increase in newspaper and Web citations and a significant decrease in book citations since 1996 (figure 1).

*Distribution of Citations*

Looking at the average of all references cited in bibliographies is useful, but, in reality, there is no "average" bibliography. By looking at the distribution of term pa-



per citations (figure 2), one can better understand the citation behavior of the class. Box and whisker plots provide a visual description of the data. On each end (the whisker) is plotted the minimum and maximum data points. In figure 2, the 95th percentile was used instead of maximum because the data often included outliers. The "box" indicates the 25th and 75th percentiles, with the median (50th percentile) represented as a horizontal bar. The box is the most informative feature because it represents 50 percent of all citations and gives an indication of the skew of the distribution.

The total number of citations in term papers has steadily increased. The median number of citations has moved from ten (in 1996) to twelve (in 1999) to thirteen (in 2000).

Book citations dropped dramatically. The median number dropped from three in 1996 to one in both 1999 and 2000. Not surprisingly, the book most often cited was the class textbook. The distribution of book citations is skewed (median near the bottom of the box), illustrating that a large proportion of 1999 and 2000 bibliographies contain no citations to books.

Journal citations dropped in 1999 bibliographies and rebounded somewhat in 2000. Although not statistically significant, the box containing 50 percent of the citations moved upward.

Newspaper citations increased from 1996. This graph illustrates very long "whiskers," indicating that some bibliog-

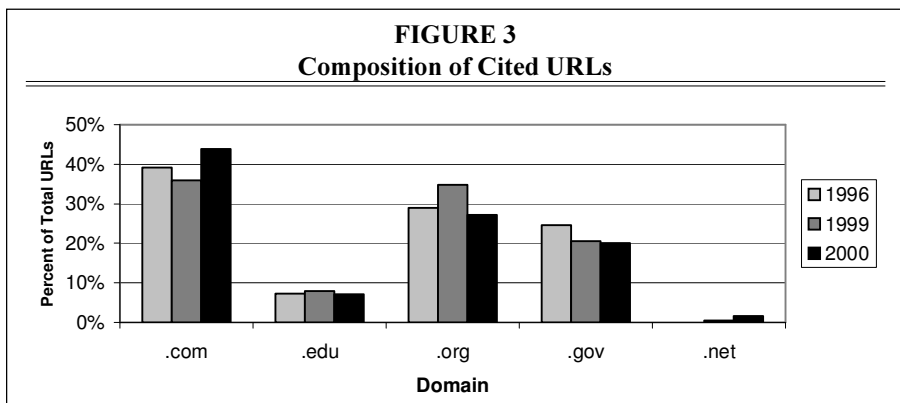
raphies (the top 75th percentile) contain an exceedingly high number of newspaper citations.

Web citations also increased from 1996. The plots for 1996 and 1999 are extremely skewed, illustrating that those who cited the Web cited large numbers of Web documents, whereas other bibliographies cited none. This phenomenon changes somewhat in 2000, as the distribution becomes more "normal" in shape. In other words, the majority of bibliographies include at least some Web sites.

To confirm a speculation that students were moving from scholarly to non-scholarly resources, book citations were combined with journal citations to form a category called "Scholarly." Newspapers and magazines also were combined to form a category called "Nonscholarly." Because many of the Web citations do not presently work, and those that do work are mutable and difficult to judge by nonexperts, Web documents were not used as part of this measure.

These scholarly citations decreased significantly ( $P < .01$ ) from 1996 to 1999 but remained virtually unchanged in 2000 bibliographies. In contrast, nonscholarly citations gradually rose from a median of three per bibliography in 1996 and 1999 to four in 2000.

In summary, the "increase in the size of bibliographies from 1999 to 2000 is fully explained by increases in traditionally nonscholarly resources" (Web sites and newspapers).



**Relationship of Citations to Grade**

Regression analysis was used to test whether the composition of bibliographies had any relationship to the grade assigned to the 2000 term papers. It was anticipated that there would be a positive relationship between the number of scholarly citations and grade or, conversely, a negative rela-

---

**The results of this 2000 update are disappointing, but not surprising.**

---

tionship between nonscholarly citations and grade.

No significant relationships were found (positive or negative) between grade and total number of citations, number of Web citations, number of scholarly citations, or number of nonscholarly citations. Although these findings do not look favorably on those grading the papers, it is important to remember that grades also reflect the quality of the writing, analysis, grammar, etc. In addition, part of the final grade was given for an oral presentation of the paper.

**Composition of Cited URLs**

Web sites became a dominantly cited genera in 2000 (figure 1), comprising 22 percent of all citations. Since 1996, there has been relatively little change in the composition of cited domains. The dot-coms continue to be the mostly heavily cited category (figure 3).

**Persistency of URLs**

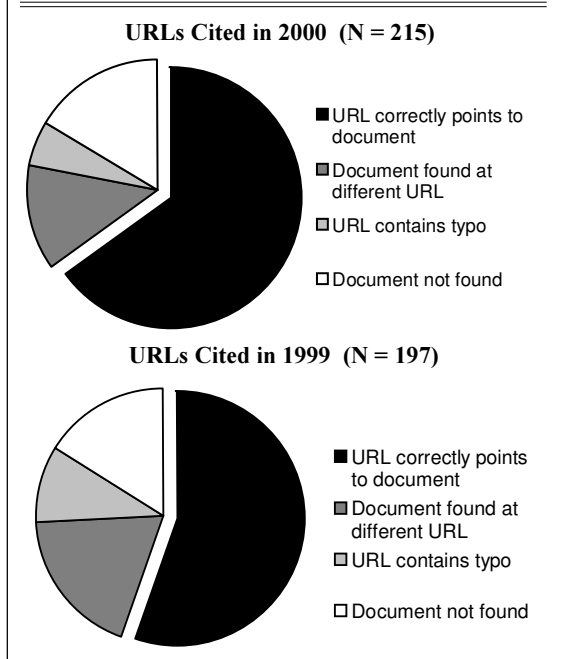
Internet citations from 1999 and 2000 bibliographies were checked for accuracy and persistency six months after submission. In 2000, the professor began requiring students to submit their papers electronically. The comparison with 1999 investigates whether electronic submission has any effect on the accuracy of cited URLs.

In 2000, 65 percent of Internet citations pointed directly to the referenced document, compared to 55 percent in 1999 (figure 4). This represents a significant change from the year before ( $P < .01$ ). Thirteen percent of cited Internet documents were found at a different URL in 2000, compared to 19 percent in 1999. Typographical errors amounted to six percent in 2000 and ten percent in 1999. In both years, 16 percent of the citations to Internet documents could not be found at all.

**Discussion**

After the results of the first study were known, the professor made it clear to his teaching assistants and librarians working with the class that he wanted to see good examples of scholarly research. Bibliographic classes taught by librarians reinforced this point. The only element that was not changed was the wording in the term paper assignment. No minimum re-

**FIGURE 4**  
**Persistency of URLs Cited in 2000 and 1999**  
**Checked after Six Months**



quirements, guidelines, or examples of scholarly resources were given.

The results of the 2000 update suggest that the professor's verbal direction had little (if any) effect on improving the scholarly component of research papers. The number of traditional scholarly materials cited this year was similar to previous years. Bibliographies grew, but only in respect to additional Web sites and newspapers. When viewed as a percentage of total citations, the "scholarliness" of bibliographies continued to decline.

Students may have relied entirely on the written assignment and ignored any verbal instructions given by TAs or librarians. The power of written expectations seems consistent with core pedagogical tenets as well as with the experiences of reference librarians. Because there was no written change in the professor's expectations, there should be no expected change in the outcome (the student papers).

Although this seems to be the most plausible explanation, several other possible explanations exist, including:

- Our unwillingness to evaluate the scholarliness of Web sites may have created unreliable results. As librarians and presumed nonexperts in the field of microeconomics, we felt they lacked the subject expertise necessary to evaluate the Web sites. It may be that many of the Internet documents cited were of scholarly value and could have been added as "scholarly" resources. This explanation is somewhat plausible, but unlikely. The majority of cited Web sites pointed to dot-com domains and, for the most part, contained news or other clearly nonscholarly resources. If a subject expert had evaluated the Web sites (those that could be found), the general breakdown might have been more heavily weighted to nonscholarly resources. In other words, the current analysis excluding Web sites provided a conservative estimate.

- Citation behavior may have no relationship to the quality of the term paper. The researchers were interested in analyzing student citations to better understand their information-searching behavior. Although a complete and concise bibliography is important to the quality of academic literature, it may have little to do with the quality of undergraduate term papers. This theory explains why no relationship was found between the bibliographies and student grades. This explanation also seems plausible but forces the researchers to accept low research standards for these students, which are reinforced by the professor's grading.

Whether or not the scholarly composition of bibliographies is of any concern to professors, it is evident that student bibliographies consist of large numbers of Web sites. This research illustrates that papers submitted electronically cite Web sites with higher accuracy and fewer typos. Most word processors work with popular Web browsers and allow users to bring up the referenced document by merely clicking on the URL. It also is easy to cut and paste the URL into a browser. In both cases, electronic submission makes it much easier for professors to verify Internet citations for accuracy and plagiarism.

### Conclusion

The results of this 2000 update are disappointing, but not surprising. A possible crisis in undergraduate scholarship is at hand, and there is no simple answer. What is clear is that librarians are not the entire solution. Professors, if they wish to see an improvement in the resources cited by students, will have to provide more clearly defined expectations in their assignments. Librarians have an opportunity to work with professors in developing research guidelines for student research assignments. This collaboration is necessary if librarians are to have any real impact on the education of students.

---

### Notes

1. Philip M. Davis and Suzanne A. Cohen, "The Effect of the Web on Undergraduate Cita-

tion Behavior 1996–1999,” *Journal of the American Society for Information Science and Technology* 52 (Feb 15, 2001): 309–14.

2. Ibid.

3. Joann E. D’Esposito and Rachel M. Gardner, “University Students’ Perceptions of the Internet: An Exploratory Study,” *Journal of Academic Librarianship* 25 (Nov 1999): 456–61.

4. Susan Davis Herring, “Faculty Acceptance of the World Wide Web for Student Research,” *College & Research Libraries* 62 (Mar. 2001): 251–58.

5. Deborah J. Grimes and Carl H. Boening, “Worries with the Web: A Look at Student Use of Web Resources,” *College & Research Libraries* 62 (Jan 2001): 11–23.

6. Patricia Senn Breivik, “Education for the Information Age,” *New Directions for Higher Education* 78 (summer 1992): 5–13.

7. ACRL, “Task Force on Academic Library Outcomes Assessment Report” (June 1998). Available online from <http://www.ala.org/acrl/outcome.html>.

8. Steve Lawrence et al., “Persistence of Web References in Scientific Research,” *Computer* 34 (Feb. 2001):26–31.