Analysis of Citations in Undergraduate Papers

Stacey Knight-Davis
Eastern Illinois University, slknight@eiu.edu

Jan S. Sung
University of Hawaii at Manoa, jansung@hawaii.edu

Follow this and additional works at: http://thekeep.eiu.edu/lib_fac
Part of the Library and Information Science Commons

Recommended Citation

This Article is brought to you for free and open access by the Booth Library at The Keep. It has been accepted for inclusion in Faculty Research & Creative Activity by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.
Analysis of Citations in Undergraduate Papers

Stacey Knight-Davis and Jan S. Sung

This paper presents the findings of a citation analysis of papers written by undergraduate students. The analysis included the types of materials cited, number of citations per paper, publication year, online availability, and refereed status of materials cited. Library ownership of materials was also analyzed. Number of citations in each paper increased over the first three papers, as did the number of refereed journals cited. There was also a positive correlation between the number of citations in the paper and the word count of the paper.

Citation analysis is a useful method to gather information about the materials students have available when writing papers. Student citations offer an opportunity to assess library collections and services. A library’s collections can be compared to works cited to see if they hold the items used most often by students. Citations can also be analyzed to determine the format, date, and refereed status of the materials cited. This kind of data is useful in determining whether or not information literacy programs are effective in encouraging the use of scholarly sources.

This project began with the desire to improve the collections and instruction services at Eastern Illinois University’s Booth Library by better understanding which sources undergraduate students were citing in their papers. We discovered that our campus had already built a depository of student papers to assess student writing. This collection included papers drawn from the entire undergraduate student body, with papers collected from the same student periodically to show changes over time. This collection of papers is unique because it represents truly authentic examples of student writing. Students were free to submit any paper they considered representative of their writing ability. As there were not strict criteria in place, and each paper was also a graded assignment, the papers submitted by the students offer a snapshot of genuine student writing. This paper collection contains accurate representations of the papers students write for course credit. The study of this excellent collection of student work provides baseline data for future instruction and information literacy programs. It also serves to collect evidence to guide collection development by describing the types of materials students cite, their use of online resources, and the availability of materials cited in the library’s collection.

Literature Review

Citation analysis has a long history as a collection development tool. Librarians have examined citations to determine the
value and need for specific materials for decades. More recently, citation analysis has become a tool librarians can use to assess the information literacy skills of students.

In 1998, Margaret J. Sylvia studied both graduate and undergraduate psychology student bibliographies to evaluate the use of her institution's journal collection. Sylvia found that 70 percent of the journals cited were held by the library. Erin T. Smith also performed citation analysis to evaluate collections, studying theses and dissertations. Books, periodicals, conference proceedings, theses, dissertations, and Web site citations were analyzed. The library owned 87 percent of all materials cited.

Citation analyses performed to assess information literacy include the landmark study by Phillip M. Davis and Suzanne Cohen with undergraduate economics students. Davis replicated his own study in 2002, finding a significant increase in the citation of Web sites between 1996 and 2000. He also found that library instruction alone was not effective in increasing the number of journal and book citations in student bibliographies. Andrew M. Robinson and Karen Schlegl studied bibliographies from political science students. Their findings showed that an effective method to raise the number of scholarly sources in student bibliographies was to combine a library instruction session with an academic penalty for using nonscholarly sources.

Other studies providing data on the types of cited materials include Joseph R. Kraus' analysis of papers presented at an undergraduate research symposium for advanced biology. Kraus found the majority (76.2%) of citations were to scholarly journals. Only 1 percent of citations were to Web sites. Charles Oppenheim and Richard Smith analyzed final-year projects from Information Science students, finding books (40.2%) and journals (29.5%) to be the most heavily cited. Fei Yu, Jane Sullivan, and Leith Woodhall analyzed bibliographies from engineering students, finding that citations to Internet sources decrease as students progress in their course of study. For first-year bibliographies, 67.1 percent of citations were to Web sites, while in bibliographies from fourth-year students 22.9 percent of citations were to Web sites.

All of the studies mentioned above use student work selected from a single discipline, course, or year of study. In contrast, Jake Carlson selected papers from classes across all disciplines that assigned a research paper. The research paper assignment required sources beyond class readings. Carlson found that the majority of sources cited (50%) were books, Journal articles were the next most commonly cited, at 19 percent, followed by Web sites at 16 percent. First-year students cited Web sites more often than journals, but all other classes used journal articles more than Web sites.

**Method**

Student work analyzed in this study comes from the Electronic Writing Portfolio. Each undergraduate student at Eastern Illinois University (EIU) is required to periodically submit writing samples to build an individual portfolio. The Electronic Writing Portfolio was developed by the EIU Committee for the Assessment of Student Learning (CASL) to assess student writing effectiveness. For submission 1, students must submit a paper written in a 1000-level English course before they have earned 30 credit hours. Submission 2 must come from a writing-centered or writing-intensive course below the 4000 level and be submitted when the student has earned between 30 and 59 credit hours. Submission 3 is made when the student has between 60 and 89 credit hours completed and must be from a 3000- or 4000-level writing-centered or writing-intensive course. Students who have completed 75 hours are required to complete a senior seminar and submit work from that course to their writing portfolio. All submission 4 papers come from senior seminar. The senior seminars...
are designed to be interdisciplinary cap-
stone courses, and students must choose a seminar topic outside their major area of study.

When the sample used in this study was drawn, students could submit any type of work, including creative and technical writing, as long as work was submitted from each required level. While students are encouraged to submit their best work, they are limited by the number and types of writing projects produced in each course that meets the requirements for each submission level. Submissions may only come from certain classes, so the documents in the portfolio become a snapshot of the types of assignments required in these classes.

Electronic writing portfolio documents are stored on a server on the EIU campus. A program to pull random samples was developed locally under the guidance of the EIU Center for Academic Support and Achievement (CASA). In 2005, a sample of completed portfolios containing work submitted between 2000 and 2005 was pulled to assess writing ability. Permission was granted by the CASA director for this sample to be used for this study also. The sample included 957 papers from 312 portfolios. The average total undergraduate student population from 2000 through 2005 was 9,690 students. Not all portfolios contained all four submissions, as transfer students were generally exempted from the first two submission requirements. The year each document was submitted was available in the sample data, but no information about the student, such as major or college, was provided.

Information on sources used in student writing was gathered by one person. Each paper was available electronically as a separate file. Each file was opened and skimmed on screen to find footnotes or a reference list. The publication year for each source found was recorded. Sources were then coded as one of the following categories: Book, Web, Journal, Newspaper, Government Document, or Audiovisual. Sources not fitting into these six categories were coded as Other. Library ownership of books was determined by searching the online catalog. Journals were checked against a list that included both print and electronic journal holdings, as well as the contents of aggregated databases accessible through the library. Journal titles were further coded as Refereed or Not Refereed. Refereed status was determined using Ulrich’s Periodicals Directory.

All identifying information on the student was removed from the paper by the CASA staff. Each paper was assigned an identifying number, and the year it was written was recorded. A word count was made for each paper using Microsoft Word.

Data were initially recorded in a spreadsheet and later converted for analysis with SPSS (Statistical Package for Social Science 14.0). Data were heavily treated using recoding, frequencies, and cross-tabulation. Statistical tests employed were Chi-Square, ANOVA, and correlation with two-tailed significance tests at the 0.05 level.

Results
From the sample of 957 papers, 293 had no citations, footnotes, or reference list of any kind. Another 244 papers included in-text citations but no reference list or footnotes (table 1). We defined in-text citation as any bibliographic information inserted into the text of the document. This included standard in-text citations in APA and MLA formats, as well as parenthetical fragments of information. We encountered page numbers without a title or author, Web page titles, author names, and URLs, all of which we defined as “in-text citations.” We defined reference list as any bibliographic information at the end of the document or as a footnote. Again, this category included standard citation formats as well as bare URLs and other citation fragments.

A total of 420 papers had reference lists that could be analyzed. The percentage of papers that had no citations of any kind remained relatively the same over the
four submissions. However, the number of papers with in-text citations that lacked a reference list decreased steadily over the four submissions. For submission 1, 42 percent lacked a reference list. Submission 2 had 28 percent with no reference list; and in submission 3, 24 percent had no reference list. For submission 4, 17 percent of papers had no reference list. The differences in citing behaviors of students over the four submissions were statistically significant (Chi-square=36.44, df=6, p<0.001).

Word counts (table 2) were significantly affected by whether students cited sources or not (F=65.05, df=2, p<0.001). While the average word count of papers without any form of citation was 946, average word count of papers with a reference list was 1,651. The correlation coefficient between word counts and the number of citations among 415 papers with a reference list was r=0.554 (p<0.001). Thirty percent (r^2=0.3) of the variation in word count can be accounted for by the number of citations. Simply put, papers with more citations will typically have a higher word count. The word count of the reference list or footnotes was included in the word count for each paper. However, the difference in word count between papers with no citations of any kind and papers with a reference list is much higher than the number of words in the reference list.

All of the following analyses were based on the 420 papers with reference lists. The information available in the papers without reference lists was too fragmentary and vague to be usable. Table 3 shows the mean number of citations, by type of source, for each submission number. The average overall citations increased over time from 3.57 for submission 1 to 5.13 for submission 3. This trend is broken at submission 4, where the average dropped to 4.88. Differences in mean number of citations among submissions were statistically significant (F=3.22, df=3, p<0.02). However, at the individual level, only the difference between

### TABLE 1
Number of Papers with/without Citations

<table>
<thead>
<tr>
<th>Papers</th>
<th>Sub 1</th>
<th>Sub 2</th>
<th>Sub 3</th>
<th>Sub 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papers without In-Text Citation or</td>
<td>38 (25%)</td>
<td>64 (33%)</td>
<td>92 (30%)</td>
<td>99 (32%)</td>
<td>293 (31%)</td>
</tr>
<tr>
<td>Reference List</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papers with In-Text Citations but</td>
<td>64 (42%)</td>
<td>54 (28%)</td>
<td>74 (24%)</td>
<td>52 (17%)</td>
<td>244 (25%)</td>
</tr>
<tr>
<td>without Reference List</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papers with In-Text Citation and</td>
<td>51 (33%)</td>
<td>74 (39%)</td>
<td>140 (46%)</td>
<td>155 (51%)</td>
<td>420 (44%)</td>
</tr>
<tr>
<td>Reference List</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>153 (100%)</td>
<td>192 (100%)</td>
<td>306 (100%)</td>
<td>306 (100%)</td>
<td>957 (100%)</td>
</tr>
</tbody>
</table>

Chi-square=36.44, df=6, p<0.001

### TABLE 2
Word Counts and Citations

<table>
<thead>
<tr>
<th>Paper</th>
<th>Word Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Papers without In-Text Citation or Reference List</td>
<td>288</td>
</tr>
<tr>
<td>Papers with In-Text Citations but without Reference List</td>
<td>243</td>
</tr>
<tr>
<td>Papers with In-Text Citation and Reference List</td>
<td>415*</td>
</tr>
<tr>
<td>Total</td>
<td>946</td>
</tr>
</tbody>
</table>

F=65.61, df=2, p<0.001

*Papers without word counts, total of 5 papers, were not included.
submission 1 and submission 3 was statistically significant.

Figure 1 provides more information on the number of citations per paper. Most students cited 5 or fewer sources in their papers. Only 37 of the 420 papers analyzed had more than 10 citations. Table 3 breaks down the citations per paper to each submission. The mean number of citations per paper for all four submissions combined was 4.42.

Figure 2 illustrates the use of different source types. The books category includes monographs, reference books, two e-
FIGURE 2
Percentage of Papers with Certain Type of Sources by Submission Number

Actual number of papers citing certain types of sources was denoted in parentheses. For example, 65% or 33 papers submitted during submission 1 cited book(s).

books, and an electronic encyclopedia. Web is defined as electronic documents that are not electronic books, journals, newspapers, or government documents. Journals include both print and electronic serial publications that are not newspapers or government documents. Magazines are included in the journals category. The other category includes annual reports, personal interviews, brochures, instruments, and works for which a source type could not be determined due to lack of information in the citation. Newspaper includes both print and electronic newspapers. The Gov Docs category also includes both print and electronic documents. Audiovisual includes films, television shows, recordings of songs, and two-dimensional artwork.

Books were the preferred source for submissions 1, 2, and 3. The percentage of papers with book sources increased from

<table>
<thead>
<tr>
<th>Number of Source Types</th>
<th>Submission 1</th>
<th>Submission 2</th>
<th>Submission 3</th>
<th>Submission 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29 (57%)</td>
<td>39 (53%)</td>
<td>72 (51%)</td>
<td>67 (43%)</td>
<td>207 (49%)</td>
</tr>
<tr>
<td>2</td>
<td>17 (33%)</td>
<td>26 (35%)</td>
<td>35 (25%)</td>
<td>49 (32%)</td>
<td>127 (30%)</td>
</tr>
<tr>
<td>3</td>
<td>5 (10%)</td>
<td>7 (9%)</td>
<td>24 (17%)</td>
<td>29 (19%)</td>
<td>65 (15%)</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2 (3%)</td>
<td>9 (7%)</td>
<td>8 (5%)</td>
<td>19 (5%)</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2 (1%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>Total</td>
<td>51 (100%)</td>
<td>74 (100%)</td>
<td>140 (100%)</td>
<td>155 (100%)</td>
<td>420 (100%)</td>
</tr>
</tbody>
</table>

Chi-Square=15.63, df=12, p<.21 (Not significant)
TABLE 5

<table>
<thead>
<tr>
<th>Online Availability</th>
<th>Submission 1</th>
<th>Submission 2</th>
<th>Submission 3</th>
<th>Submission 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Available Online</td>
<td>18 (50%)</td>
<td>33 (41%)</td>
<td>63 (23%)</td>
<td>24 (17%)</td>
<td>138 (26%)</td>
</tr>
<tr>
<td>Available Online</td>
<td>18 (50%)</td>
<td>48 (59%)</td>
<td>215 (77%)</td>
<td>115 (83%)</td>
<td>396 (74%)</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>81</td>
<td>278</td>
<td>139</td>
<td>534</td>
</tr>
</tbody>
</table>

Chi-Square=27.15, df=3, p<0.001

Table 4 shows the number of papers with citations to a single source type and the number of papers with multiple source types in the same paper. Of the 420 papers with a reference list, nearly 50 percent of the papers cited only a single source type. Among 207 papers with a single source type, almost half of them, or 99 papers, cited only books. Web sources were the only source for 46 papers, and 38 papers cited only journal sources. The number of papers with a single source type decreased over time from 57 percent for submission 1 to 43 percent for submission 4. The use of multiple source types increased over time, but the difference did not reach statistical significance (Chi-Square=15.63, df=12, p<0.21).

For all source types combined, 57 percent of citations were to materials available online. These materials included Web sites as well as newspapers (60 online/79 total), books (5 online/587 total), govern-
ment documents (40 online/60 total), and journals available on the Internet (396 online/534 total). For journals, 396 articles, or 74 percent out of 534 articles cited, were available online. Online availability of articles increased steadily, from 50 percent for submission 1 to 83 percent for the last submission from the senior seminar (table 5). It is not known if students used the online version of these materials or the print version. It is only known that the online version was available to them.

Table 6 shows the pattern of publication years and source types. For all source types, students tended to use more current materials. More than 65 percent of books and 85 percent of articles cited were published after 1991.

Overall, students cited refereed journals more often than nonrefereed, with 61 percent of journal citations to refereed titles (figure 3). The use of refereed articles doubled from 31 percent for submission 1 to 63 percent for submission 2. The percentage increased again to 77 percent for submission 3 but decreased to 36 percent for submission 4. This difference was statistically significant (Chi-Square=80.63, df=3, $p<0.001$).

Figure 4 and figure 5 show whether the library held the books and articles cited by the students. Overall, 55 percent of cited books were held by the library. Many of the cited books not owned by the library were course textbooks, which the library does not generally include in its collection because all students are issued textbooks from the university textbook rental service. Journals, which are typically available only through a library, are a better indicator of the match between library holdings and student use. The library offered access to 80 percent of the articles students cited.

Discussion
It is clear that the submission number of the paper, which is a reflection of the course taken and the amount of time the student has been in college, is related to the type and amount of material that students cite. In general, we had expected that the number of reference list citations per paper would increase from submission 1 to submission 4, but this was not the case. While it is difficult to completely establish which factors most heavily influence student citation behavior, some trends are clear.

The number of papers that included a reference list did increase over the four submissions, as expected. For submission
1, only 33 percent of papers had a reference list, but submission 4 had a reference list in more than 51 percent of the papers. This aspect of citation behavior has a simple relationship to the amount of time the student has spent in school. The more hours a student has completed, the more likely he or she is to cite sources.

The large number of papers without citations was disappointing when first discovered. This result was partially explained by the CASA staff. Students are encouraged to submit their best writing, not their best research. Many students choose to submit essays, creative writing, or reflective pieces. CASA staff also ex-
plained that some students choose to keep their works-cited list as a separate file. Students are not required to combine the works-cited file with the file containing the body of the paper before submitting, which resulted in a group of papers with in-text citations but no reference list.

In addition to the positive relationship between including a reference list and number of class hours completed, the analyses performed here also show a positive relationship between paper length and word count. While the idea that longer papers would cite more sources is intuitive, it is reassuring to see that this idea holds up with a large sample of papers from many disciplines. Averaged across all submissions, students who cited sources and had a reference list wrote 705 more words than those who did not cite sources and did not include a reference list. The reference lists were included in the word count. It is estimated that the typical citation in a reference list for papers in this study was 19 words.

As noted in other studies, books were the preferred source for many students. For all submissions combined, books accounted for 59 percent of citations. This is comparable to Carlson’s results of 50 percent of citations to book sources. However, the mean number of citations per paper found by Carlson was much higher. In this study, mean number of citations for submission 1 was 3.57. The mean for submission 4 was 4.88. In comparison, Carlson’s first-year student papers had a mean of 7.78 citations per paper, and senior papers had a mean of 10.11. This difference is most likely due to Carlson’s sampling method, which selected papers only from classes assigning a research paper.

The number of citations to journals was highest for submission 3, with a mean of 1.99. This is lower than Carlson’s finding of a mean of 2.57 citations to journals in papers written for 3000- and 4000-level courses. However, Carlson found that the mean number of citations to journals in papers by juniors in the humanities was only 0.43. As the exact proportions for the number of students in each discipline was not known in this study, it is not possible to determine whether the relatively low number of journal citations in 3000- and 4000-level courses is due to discipline-specific citation practices or some other factor.

Use of refereed journals increased sharply for submission 3, with 77 percent of article citations to refereed journals. Since the courses from which students can submit submission 3 typically require a research paper, the sharp increase in use of refereed articles is likely due to requirements in the paper assignment. This assumption is borne out by the sharp decrease in journal usage for submission 4.

The citation patterns seen in submission 4 papers do not follow the trends established in the first three submissions (graph 2). Journal citations drop off (28% for submission 4, 42% for submission 3), and citations to Web sites rise (61% for submission 4, 38% for submission 3). Citations to audiovisual materials also rise (13% for submission 4, 3% for submission 3). The number of citations to nonrefereed journals increased as well (64% for submission 4, 23% for submission 3).

All papers for submission 4 come from senior seminar classes. Two circumstantial factors were identified that may contribute to the citation patterns seen in papers from these classes. First, a study of writing effectiveness conducted by CASA on this same sample found that senior papers were of lesser quality than papers from other classes. The second factor is the lack of a coordinated library instruction program for senior seminar students. Robinson and Schlegl found that instruction together with enforced instructor guidelines are an effective way of generating quality student bibliographies.

Considering these two factors, we suspect that the lack of standardization in writing requirements for the senior seminar classes is responsible for the citation pattern seen in submission 4 papers. The goal of all senior seminar courses is
to provide a “cross-disciplinary culminating experience that will provide students with an opportunity to apply concepts and use skills developed in both their general education and major courses,” and all courses must include a writing component. There are requirements on the percentage of the grade that must be based on writing and also a requirement that at least one paper must be revised, but there is no requirement on the length of those papers.

The wide range of topics and teaching styles used in these courses leads to a huge variability in the assignments given, and the lack of standards for the length of writing assignments allows for a large variation in the length of papers submitted. While all courses must address literacy, critical thinking, and citizenship, library research is not specifically required. Assessment of these courses is carried out by the instructor, based on his or her set learning objectives.

It is possible to argue that students simply revert to more basic Web searching and more basic magazine articles when faced with material from outside their major in senior seminar. However, the poor writing quality found in these courses suggests a problem with the assignment or poor revision. Since only one assignment must be revised, it is possible that students submit assignments that did not require revision.

Conclusion
This analysis of student citation behavior reaffirmed some of the assumptions we had made about our collections. It also gave us further insight into the types of materials students cite and how many citations their papers typically include. Having this study completed allows Booth Library to make decisions based on the evidence of past student behavior instead of anecdotal observances and assumptions.

One of the assumptions confirmed by this study is the heavy use of materials that are available on the Internet. Over all four submissions, 74 percent of the articles cited were available online. For submission 4, 83 percent of articles cited were available online. We are pleased to note that eleven journals cited by students in print were later licensed in electronic form. Having evidence of undergraduate students’ strong preference for electronic content is helpful when discussing electronic-only journal subscriptions with the teaching faculty. The large percentage of online material cited in student papers also provides support for licensing electronic content.

In addition to providing baseline data on citation behavior, this project also provided important insight into current library instruction programs at our university. At this time, library instruction is offered to any instructor requesting it. Each instructor determines type and amount of library instruction students receive. A dialogue has begun between the library and academic departments to determine if the current system is meeting the needs of students and instructors. Should changes be made in curriculum or in the way library instruction is offered, this study will provide valuable baseline data to assess the effectiveness of the change.

An important point to be taken from this study is the importance of the requirements of the paper assignment on citation behavior. The sources a student will choose to cite are often heavily influenced by the requirements regarding sources, or lack thereof, in the paper assignment. Librarians typically see only students who are actively seeking sources and may overlook students who have not had classes that required them to conduct research or cite sources. It is important to remember this diversity of student experience when conducting bibliographic instruction sessions and helping students in the reference room.

Another benefit from this project was that it allowed Booth Library to develop a good working relationship with the people and committees responsible for student
assessment on our campus. The library has also become more involved in campus assessment activities. A library faculty member now belongs to the Committee for the Assessment of Student Learning (CASL). We hope that this involvement will ensure that the library’s assessment needs are considered as changes to the Electronic Writing Portfolio procedures are discussed. A major limitation of this study was that no information on academic discipline or college was available. We hope that, by having a library representative at CASL, we can encourage changes that permit this data to be collected.

While this study describes the citation behavior of students at one university, all libraries should take an interest in what students at their institutions are writing and how well the library is supporting student needs. While citation analysis can show us what students choose to cite, additional research is needed to determine why students choose these sources. It is imperative that all librarians and information professionals continually examine the behavior of their users and clients to ensure that the best possible services and resources are provided.

Notes

10. Ibid.
11. Ibid.