

# Citation Autobiography: An Investigation of ISI Database Coverage in Determining Author Citedness

Thomas E. Nisonger

This article presents a case study investigating the coverage completeness of the Institute for Scientific Information's citation data for specific authors, based on analysis of this author's lifetime citation record, which was compiled through the ISI database, searching the literature for nearly fifteen years, and through various Web search engines. It was found that (with self-citations disregarded) the ISI captured 28.8 percent of the total citations, 42.2 percent of print citations, 20.3 percent of citations from outside the United States, and 2.3 percent of non-English citations. The definition and classification of Web citations are discussed. It is suggested that librarians and faculty should not rely solely on ISI author citation counts, especially when demonstration of international impact is important.



Most academic librarians are undoubtedly familiar with the Institute for Scientific Information (ISI) citation databases, which can be used for innumerable reference and evaluation purposes. The *Science Citation Index*, the *Social Sciences Citation Index*, and the *Arts & Humanities Citation Index* were developed by ISI during the late 1960s and early 1970s. These tools were issued quarterly and cumulated annually. Among other uses, they provided—based on the several thousand journals in the ISI database—a listing of an author's journal publications during the current year (through the Source Index) and a list of the citations received that year by any work of the author regardless of when published (through the

Citation Index). However, citations to multiple-authored works could be retrieved only under the first author's name.

In the later half of the 1990s, ISI began marketing the *Web of Science* (updated continuously), which provides a Web-based interface to the three ISI citation indexes in which one can retrieve an author's cumulative citation and publication record since 1987 based on ISI source journals. The more recently introduced *Web of Knowledge* incorporates the *Web of Science*, *Current Contents*, and the *Journal Citation Reports*.

The capacity to access author citation records can support numerous evaluative functions, including the promotion and tenure process in universities. As Kathryn

---

*Thomas E. Nisonger is an Associate Professor in the Indiana University School of Library and Information Science, Bloomington; e-mail: nisonge@indiana.edu. The author thanks his two graduate assistants at Indiana University's School of Library and Information Science, Jason Cooper and Leah Broadadda, who helped tabulate data; his wife, Claire Nisonger, who helped search for Web citations; and Debora Shaw, Interim Dean of the School, who kindly shared one of her unpublished manuscripts with him.*

L. Reed wrote, "Citation analysis has become important to faculty members being reviewed for promotion or tenure."<sup>1</sup> Indeed, there is considerable evidence that promotion and tenure candidates often include their lifetime citation records as part of their dossier. ISI data regarding author citedness frequently has been used to identify the most-cited authors in an academic field and to rank or evaluate universities, academic departments, and research laboratories based on citations received by their faculty. (See the literature review section for some specific examples in the field of library and information science [LIS]).

Yet, as many writers, including Edward Truman Funkhouser, Barbara A. Rice and Tony Stankus, and Reed, have pointed out, ISI citation data are incomplete.<sup>2-4</sup> They are limited to citations from the approximately 6,500 source journals in the ISI database. Citations from non-ISI source journals, books, and Web documents as well as from other formats are not counted. Although the *Web of Science* offers relatively quick and convenient access to an author's citation record, it is an incomplete record.

This research project uses an intensive analysis of this author's own lifetime citation record as a case study to investigate the parameters of ISI's coverage for a LIS researcher. Although the author's citation count is far behind the most highly cited LIS researchers, his lifetime total of more than three hundred citations from print sources and more than four hundred with the inclusion of Web citations is not trivial and provides a reasonably robust size data set for analysis. He ranked among the top-third of Indiana University School of Library and Information faculty, according to lifetime citations received in the three ISI citation indexes, in Blaise Cronin and Kara Overfelt's study.<sup>5</sup> The term "citation autobiography" is introduced because the author has compiled his own citation record. This exercise was undertaken initially to support the author's case in the promotion and tenure process but then

gained a "life of its own" as the author continued the process after tenure was granted. It gradually became apparent that the project offered a unique case study that could serve as an example to others in the field because few individuals would spend years of labor-intensive effort meticulously compiling a lifetime citation record.

### **The Role of Citation Data in the Promotion and Tenure Process**

One of the primary purposes of a promotion and tenure dossier is to demonstrate the quality, significance, and impact of the candidate's research. Among the appropriate evaluative criteria, Pamela S. Bradigan and Carol A. Mularski listed citation data, along with publication count, the refereed status plus general stature of the journals in which the candidate had published, reviews of books by the candidate, OCLC holdings counts for his or her books, awards for outstanding research, and assessment letters from internal and external reviewers.<sup>6</sup> The fact that a promotion and tenure candidate's work has been cited in another publication demonstrates that it has contributed to the scholarly communication process and is generally considered evidence of the research's quality, visibility, or impact. Yet, citation data are controversial. Critics point out that citations can be negative and that citation counts can be influenced by such factors as the number of authors in a field, variant citation practices among disciplines, and writing on a "hot topic."<sup>7</sup>

Rather than merely counting citations, a promotion and tenure dossier may quote positive comments, include photocopies of the citing documents' segments where the candidate was mentioned, or analyze the citations according to specific variables such as country of origin (as citations from outside the United States might be presented as evidence of international impact). Although beyond this article's scope, another application of citation data in the promotion and tenure evaluation process concerns whether the

candidate has published in high- or low-impact-factor journals as indicated in the *Journal Citation Reports* (available through the *Web of Knowledge*).<sup>8</sup> Thus, it is acknowledged that citation data constitute simply one of numerous methods to demonstrate research impact—and even then they are somewhat controversial. Moreover, an author's raw citation count is only one type of citation data.

Nevertheless, candidates wishing to analyze their citations in-depth to help substantiate their work's impact must first identify the set of citations upon which to base their analysis. It is generally known that an author can fairly easily identify the citations to his or her work that are included in the ISI databases, but that the record is incomplete. By examining the extent of ISI's coverage, this case study is potentially useful to librarians and scholars trying to decide whether an author citation search should be limited to that database or extended beyond it. The findings also have implications for evaluation studies based on authorship citation data.

### Literature Review

This literature review mentions salient publications about the citation records of specific authors, the use of citation data for ranking LIS researchers and education programs, and the evaluation of the *Web of Science* as well as ISI database coverage. The phrase "ego-centered citation analysis" was applied by Howard D. White to the use of a specific author as a starting point for mapping the citation relationship among an author network, although he explained that the term does not connote ego-centricity or egotism.<sup>9</sup> In their citation analysis of Blaise Cronin, Stephen B. Harter, and Rob Kling, Cronin and Debora Shaw observed that bibliometric techniques can provide "biographical sketches of authors," a phrase they attribute to Christine L. Borgman and Jonathan Furner.<sup>10,11</sup> The term "personal citation index" is applied by Bernie Sloan to his record of references to his work in print and on the Web.<sup>12</sup> Heting

Chu analyzed 377 citing documents to the work of her doctoral advisor, the illustrious information scientist Belver C. Griffith.<sup>13</sup> Michael J. Moravcsik applied a citation classification system to ten years' worth of citations to the article he had coauthored with Poovanalingam Murugesan proposing the system.<sup>14,15</sup>

Robert M. Hayes used normalized citation data from the Social Sciences Citation Index covering 1965 to 1980 to identify in rank order the top forty North American LIS researchers and the top twenty North American LIS education programs.<sup>16</sup> Partially replicating Hayes, John M. Budd and Charles A. Seavey employed 1981 to 1992 *Social Sciences Citation Index* data to rank the thirty-three most-cited LIS researchers and the leading twenty North American LIS education programs according to several criteria, including the total number of citations received by their faculty and the number of citations prorated for faculty size. Incorporating numerous variables, they offered a final ranking of the top fifteen programs.<sup>17</sup> Budd later updated the study with 1993–1998 *Social Sciences Citation Index* data.<sup>18</sup> It should be noted that similar-type rankings done in disciplines other than LIS are beyond this review's scope.

The *Web of Science* has been reviewed by Deborah Lynne Wiley and Victor Rosenberg, whereas the *Web of Knowledge* has been reviewed by Barbara E. Quint.<sup>19–21</sup> Funkhouser analyzed the references in twenty-seven communications journals (thirteen covered by ISI and fourteen not covered) during 1990 and found that 26 percent of author citations would have been lost because they were from non-ISI journals. Moreover, twenty-seven of the fifty most highly cited authors received at least 25 percent of their citations from non-ISI journals.<sup>22</sup> G. Van Hooydonk and Greta Milis-Proost calculated that the ISI captured 16 percent of the citations and 18 percent of the cited publications of 258 University of Ghent professors appointed between 1817 and 1913.<sup>23</sup> On a tangential issue, a number of researchers, including

Pam Royle and Ray Over as well as Selden Durgom, have investigated the completeness of ISI publication data for faculty and other authors.<sup>24, 25</sup> (This question is not analyzed here because faculty and librarians, when preparing promotion and tenure dossiers, would obviously be aware of their complete publication record, unlike their citation record.)

## Methodology

This section explains the methodology used to identify citations, discusses the methodological issues regarding the definition and classification of Web citations, and describes the data analysis.

### Identification of Citations

The following methodology was employed:

1. *Identification of the author's citation record as recorded by ISI:* This was done through author searches in the *Web of Science*, using both the easy- and full-search options and covering 1987 through the present, and in the print or CD-ROM versions of ISI's three citation indexes—the *Social Sciences Citation Index*, the *Science Citation Index*, and the *Arts & Humanities Citation Index*—from 1980 through 1986. Note that the author's first publication was in 1980. Searches were done under three variant name forms: the author's surname plus first name initial; the author's surname plus middle two initials; and the author's surname plus first initial followed by a wild card (which identified a citation with an incorrect middle initial). The names of first-listed coauthors also were searched.<sup>26</sup> Modeled on the ISI definition, a citation was defined as an item-to-item link. Thus, three of the author's publications cited in an article counted as three citations, whereas one publication cited three times counted as simply one citation.

2. *Identification of citations to the author's works not covered by the ISI database:* The following techniques have been used at various times during the previous fifteen years as the author has meticulously compiled his lifetime citation record:

- a. reviewing all the references in more than 1,200 items included in two major monographic bibliographies in his specialty area, library collection evaluation including use of citation analysis and journal ranking<sup>27, 28</sup> as well as hundreds of other items considered for inclusion in these works;

- b. reviewing the citations in all issues published since 1980 of selected key journals in the author's specialty not covered in the ISI database;

- c. scanning the citations and bibliographies in textbooks and monographs pertinent to the author's research areas;

- d. recording of citations discovered through research and teaching activity plus professional reading throughout his career;

- e. maintaining a continuously updated file of citations as they are discovered (note that tactics b, c, d, and e are recommended by Reed<sup>29</sup>);

- f. searching the author's name in three Web search engines (e.g., Google, AltaVista, and Teoma). Seven permutations of the name were used, variously combining direct and indirect name order, full forms and initials, and formal and informal versions of the author's first name.

References to the author in award announcements, vendor advertising, book review excerpts in academic journals, and in-house publications by his university or academic program were not counted as citations. This investigation analyzes citations received through a cutoff date of February 1, 2003. Any citations or book reviews appearing in the *Web of Science* or identified through other means after that date were not included in this study.

### Definition and Classification of Web Citations

The definition, count, and classification of Web citations presented a number of methodological challenges. Cronin and others offered an eleven-component typology for classification of Web citations: abstract, article, conference, current awareness, external home page, listserv,

personal home page, resource guide, book review, syllabus, and table of contents.<sup>30</sup> Liwen Vaughan and Shaw have classified Web citations into seven categories: journal, author (i.e., a personal or departmental Web page), a Web bibliographic service, a class bibliography or reading list, citation in a paper posted on the Web, conference, or "other."<sup>31</sup>

For this autobiography's purpose, a Web citation was operationally defined as a decision to reference the author's work in a document or page mounted on the Web. Library accession lists, vendor advertising, current awareness services, and table of contents listings were not counted as Web citations because they do not meet this criterion and would not be considered citations by ISI. Likewise, a simple mention of the author's name without citing a specific work, such as in a conference program or resume, was not considered a Web citation.

Five categories of Web citations were defined for tabulation in this study, as follows:

- *Manuscripts*: Included here would be preprints, research reports, student papers, and other scholarly manuscripts mounted on the Web that contain references to the author's work
- *Course syllabi or reading lists*: This fairly self-evident category consists of syllabi for university courses or reading lists for qualifying examinations, identified through Web search engines, that list publications by the author.
- *Web bibliographies*: A Web bibliography was defined as one compiled on a unified theme, such as journal evaluation or library collection management. As noted above, current awareness services for recent publications, listings of library holdings, and DBLP (Digital Bibliography and Library Project) were not counted as bibliographies. An item by the author listed in multiple versions of the *Scholarly Electronic Publishing Bibliography* by Charles W. Bailey Jr. (version 46, issued in December, 2002 and the most current one as of this research project's cutoff date) was counted only once.

- *Conference presentations*: This category comprises citations in documents presented at formal conferences or ad hoc workshops that did not appear in print form (if available in print format, the citation was considered a print rather than a Web citation).

- *Electronic journals*: This category refers to electronic-only journals identified through Web search engines. As with the preceding category, it was not considered a Web citation if a print version of the journal exists.

This scheme represents an operational taxonomy for this investigation's purposes. From a general bibliometric perspective, further research is needed concerning the definition and classification of Web citations.

### **Data Analysis**

The language analysis was based on the language of the citing document rather than the citing publication (i.e., a citation from a French-language article in an English-French bilingual journal would be counted as French) and could easily be determined by direct examination of the document itself. The country of origin for citations in print documents was determined by the publisher's location. For Web citations, the country of origin was determined by the location of the server housing the citing document. Citations from citing documents identified in dual formats were counted only once, based on the format in which the item was originally issued.

### **Results**

Table 1 summarizes the print citations to the author's work contained in the ISI databases (all the author's citations in ISI were from print documents) and non-ISI citations identified through other sources, representing a cumulative lifetime total through February 1, 2003. Book reviews are excluded. It is noteworthy that ISI captured 44.6 percent of the author's 377 print citations, 54.9 percent of 71 self-citations, and 42.2 percent of 306 citations from others. Although not stated explic-

**TABLE 1**  
**Print Citation Totals**

	ISI Citations	Non-ISI Citations	Total Citations	Percent Captured by ISI
Citations from				
Others	129	177	306	42.2%
Self-citations	39	32	71	54.9%
<b>Total Citations</b>	<b>168</b>	<b>209</b>	<b>377</b>	<b>44.6%</b>

itly in the table, 23.2 percent of the ISI citations included in ISI were self-citations, contrasted with a 15.3 percent self-citation rate for non-ISI citations and an 18.8 percent overall self-citation rate.

After tabulation of the print citations, 127 Web citations from other individuals or corporate entities were counted in accordance with the procedure outlined in the methodology section of this article. No attempt was made to identify or count Web self-citations. Accordingly, ISI captured 33.3 percent of 504 total citations and 29.8 percent of 433 citations when self-citations are disregarded.

Book reviews are tabulated separately because they are covered by ISI but are analytically distinct from citations. The *Web of Science* lists fifteen book reviews of publications by the author—two books and an

edited journal issue that received reviews as if it were a book. (A third book, coauthored in the mid-1980s, received only one review and it was not covered by ISI.) These four publications received six-

teen book reviews that fell beyond ISI's net, so the overall book review coverage stood at 48.4 percent. In addition, the author identified fourteen reviews of four other books to which he had contributed a chapter that explicitly mentioned his name in the review's text. None of these was retrieved under a search of the author's name in the *Web of Science* because he was not the book's editor, but eight (57.1%) were retrieved by searching the editor's name.

Because self-citations and book reviews often are discounted from author citation counts, the remaining analysis will disregard book reviews and the seventy-one print self-citations and focus exclusively on the 433 citations received from others. Table 2 breaks down the remaining 306 print citations by format of the citing document.

**TABLE 2**  
**Analysis of Print Citations by Format of Citing Document**  
**(Self-Citations Disregarded)**

Format	ISI	Non-ISI	Total	Percent Captured by ISI
Journal	117 (90.7%)	72 (40.7%)	189 (61.8%)	61.9%
Conference Proceeding	6 (4.7%)	5 (2.8%)	11 (3.6%)	54.5%
Annual/Series	6 (4.7%)	19 (10.7%)	25 (8.2%)	24.0%
Book	0 (0%)	45 (25.4%)	45 (14.7%)	0%
Book Chapter*	0 (0%)	12 (6.8%)	12 (3.9%)	0%
Manual/Handbook	0 (0%)	8 (4.5%)	8 (2.6%)	0%
Pamphlet	0 (0%)	10 (5.6%)	10 (3.3%)	0%
Ph.D. Dissertation	0 (0%)	4 (2.3%)	4 (1.3%)	0%
Other	0 (0%)	2** (1.1%)	2** (0.7%)	0%
<b>Total</b>	<b>129 (100.1%)</b>	<b>177 (99.9%)</b>	<b>306 (100.1%)</b>	<b>42.2%</b>

\*Throughout this article, a book chapter is defined as a contributed chapter to an edited book.

\*\* One master's paper and one library annual report.

ISI captured 61.9 percent of the citations from journals. Worded differently, 61.9 percent of the print journal citations originated from ISI source journals compared to 38.1 percent from nonsource journals. ISI also captured 54.5 percent of the citations from conference proceedings (those contained in the proceedings of American Society for Information Science, now called the American Society for Information Science and Technology) and 24 percent of the citations from annual volumes or ongoing monographic series (in this instance, the *Annual Review of Information Science and Technology*). As would be expected, none of the citations from books, book chapters, manuals, pamphlets, or dissertations was covered by ISI because these formats are not included in its source database. It is especially noteworthy that 45.8 percent of the citations not captured by ISI (presented in the column labeled "non-ISI") and 26.5 percent of the total print citations were from formats that are not even included in ISI's database.

Table 3 displays the distribution of Web citations according to the formats explained in the methodology section. Be-

**TABLE 3**  
**Distribution of Web Citations by Format**

Format	Number	Percent
Manuscript	50	39.4%
Course Syllabus	27	21.3%
Bibliography	23	18.1%
Conference Proceeding	14	11.0%
Electronic Journal	13	10.2%
<b>Total</b>	<b>127</b>	<b>100.0%</b>

cause manuscripts, course syllabi, and bibliographies *per se* are not part of its source database, ISI coverage of these categories was nonexistent. When Web citations from e-journals and conference proceedings are combined with the print figures, ISI's total coverage of journal citations declines to 57.9 percent (117 of 202) and conference proceedings to 24% (6 of 25). Note that Web citations have been included in this study in order to present a complete citation record for the author.

Table 4's analysis by the citing document's language reveals an overwhelmingly English-language profile for print citations. Yet, nearly 30 percent of

**TABLE 4**  
**Analysis of Citations Received by Language of Citing Document (Self-Citations Disregarded)**

Language	ISI Print		Non-ISI Print		Web		Non-ISI Total		Grand Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
English	128	99.2%	173	97.7%	89	70.1%	262	86.2%	390	90.1%
Spanish	1	0.8%	0	0%	7	5.5%	7	2.3%	8	1.8%
Swedish	0	0%	0	0%	8	6.3%	8	2.6%	8	1.8%
German	0	0%	2	1.1%	5	3.9%	7	2.3%	7	1.6%
Chinese	0	0%	2	1.1%	3	2.4%	5	1.6%	5	1.2%
Italian	0	0%	0	0%	3	2.4%	3	1.0%	3	0.7%
French	0	0%	0	0%	2	1.6%	2	0.7%	2	0.5%
Polish	0	0%	0	0%	2	1.6%	2	0.7%	2	0.5%
Portuguese	0	0%	0	0%	2	1.6%	2	0.7%	2	0.5%
Other	0	0%	0	0%	6*	4.7%	6*	2.0%	6*	1.4%
<b>Total</b>	<b>129</b>	<b>100.0%</b>	<b>177</b>	<b>99.9%</b>	<b>127</b>	<b>100.1%</b>	<b>304</b>	<b>100.1%</b>	<b>433</b>	<b>100.1%</b>

\*There was one Web citation from each of six languages: Greek, Hebrew, Hungarian, Japanese, Russian, and Turkish.

**TABLE 5**  
**Analysis of Citations Received by Citing Document's Country of Origin**  
**(Self-Citations Disregarded)**

Country of Origin	ISI Print		Non-ISI Print		Web		Non-ISI Total		Grand Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
United States	103	79.8%	143	80.8%	59	46.5%	202	66.4%	305	70.4%
United Kingdom	10	7.8%	9	5.1%	0	0%	9	3.0%	19	4.4%
Netherlands	12	9.3%	1	0.6%	3	2.4%	4	1.3%	16	3.7%
Australia	0	0%	12	6.8%	2	1.6%	14	4.6%	14	3.2%
Germany	2	1.6%	4	2.3%	5	3.9%	9	3.0%	11	2.5%
Sweden	0	0%	0	0%	11	8.7%	11	3.6%	11	2.5%
Canada	0	0%	1	0.6%	8	6.3%	9	3.0%	9	2.0%
Taiwan	0	0%	4	2.3%	3	2.4%	7	2.3%	7	1.6%
Spain	0	0%	0	0%	6	4.7%	6	2.0%	6	1.4%
Finland	0	0%	0	0%	4	3.1%	4	1.3%	4	0.9%
Italy	1	0.8%	0	0%	3	2.4%	3	1.0%	4	0.9%
Brazil	0	0%	0	0%	2	1.6%	2	0.7%	2	0.5%
France	0	0%	0	0%	2	1.6%	2	0.7%	2	0.5%
Japan	0	0%	0	0%	2	1.6%	2	0.7%	2	0.5%
Pakistan	0	0%	1	0.6%	1	0.8%	2	0.7%	2	0.5%
Poland	0	0%	0	0%	2	1.6%	2	0.7%	2	0.5%
Chile	1	0.8%	0	0%	0	0%	0	0%	1	0.2%
Switzerland	0	0%	1	0.6%	0	0%	1	0.3%	1	0.2%
Other Countries	0	0%	0	0%	10*	7.9%	10*	3.3%	10	2.3%
Unknown	0	0%	1	0.6%	4	3.1%	5	1.6%	5	1.2%
<b>Total</b>	<b>129</b>	<b>100.1%</b>	<b>177</b>	<b>100.3%</b>	<b>127</b>	<b>100.2%</b>	<b>304</b>	<b>100.2%</b>	<b>433</b>	<b>99.9%</b>

\*There was one Web citation from each of ten other countries: Argentina, Belgium, Denmark, Greece, Hungary, Iceland, Israel, Mexico, Russia, and Turkey.

the Web citations were not in English, with fourteen other languages represented. After English, 6.3 percent of the Web citations were in Swedish and 5.5 percent were in Spanish. The fact that the ISI database retrieved 32.8 percent of the English-language citations (128 of 390), but only 2.3 percent of the non-English ones (1 of 43), could arguably support the well-known perception of a pro-English bias by ISI.

Table 5 breaks down the author's citations by the citing document's country of origin. Comparison with table 4 demonstrates a broader dispersion by country of origin than by language. Twenty-eight

countries are represented in table 5 with about 70 percent of the total citations concentrated in a single nation (the United States), whereas only fifteen languages are listed in table 4 with more than 90 percent of the citations concentrated in English. This pattern is partially explained by the fact that several countries, including some non-English-speaking nations, publish or mount English-language documents on the Web. Likewise, Web citations are much more broadly dispersed by both language and country of origin than are print citations in general or the set of citations captured by ISI. Finally, ISI was most effective at retrieving



citations from the Netherlands at a 75 percent overall success rate (12 of 16), followed by the United Kingdom at 52.6 percent (10 of 19). The rate for the United States was 33.8 percent (103 of 305) and 4.3 percent for countries other than the Netherlands, the United States, or the United Kingdom (4 of 93). For all non-U.S. countries the retrieval rate was 20.6 percent (26 of 126).

The preceding analysis has focused on the citing documents; attention now should be turned to the cited documents. Because the author has published almost exclusively in English in the United States, there is no need to analyze language and country of origin variables. The format of the cited documents is presented in table 6. Note that the refereed status of journals is based on the specific article published rather than the journal title.

Most of the citations captured by ISI (64.3%) and print sources in general (71.2%) were to refereed journal articles, whereas only 29.9 percent of Web citations were to refereed articles, a proportion only slightly higher than the 29.1 percent of Web citations to nonrefereed journal articles. The explanation for this phenomenon is unclear, but one is tempted to speculate that it may partially reflect the fact that the Web itself is unrefereed and

thus the choice of documents for citation is less rigorous. The analysis shown in table 2 of the citing documents format does not distinguish between refereed and unrefereed journal articles because this information was unknown to the author. However, in a personal citation autobiography, the author obviously knows the refereed status of each cited document.

Finally, the cited documents should be examined briefly. A total of forty-five published works by the author were cited, comprising eighteen refereed journal articles, fourteen unrefereed journal articles, three book chapters, three books, three published conference proceedings, two edited journal issues, one book review, and one dissertation. The ISI databases picked up at least one citation to twenty-eight of these (62.2%) and the Web cited twenty-nine (64.4%). Of the eighteen cited refereed articles, 83.3 percent (15) were captured by ISI and 55.6 percent (10) by the Web. In contrast, ISI picked up 35.7 percent (5) of the fourteen unrefereed articles that were cited and the Web 78.6 percent (11), thus reinforcing the finding that unrefereed articles are more likely to be cited on the Web.

Ten of the forty-five cited documents (22.2%) were published prior to 1990. These ten documents accounted for 37.0

**TABLE 6**  
**Analysis of Citations Received by Cited Document's Format**  
**(Self-Citations Disregarded)**

Format	ISI Print		Non-ISI Print		Web		Non-ISI Total		Grand Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Refereed Journal	83	64.3%	126	71.2%	38	29.9%	164	53.9%	247	57.0%
Nonrefereed Journal	11	8.5%	24	13.6%	37	29.1%	61	20.1%	72	16.6%
Book	17	13.2%	20	11.3%	35	27.6%	55	18.1%	72	16.6%
Book Chapter	7	5.4%	4	2.3%	4	3.1%	8	2.6%	15	3.5%
Conference Proceed.	10	7.8%	1	0.6%	4	3.1%	5	1.6%	15	3.5%
Edited Journal Issue	0	0%	1	0.6%	8	6.3%	9	3.0%	9	2.1%
Ph.D. Dissertation	0	0%	1	0.6%	1	0.8%	2	0.7%	2	0.5%
Book Review	1	0.8%	0	0%	0	0%	0	0%	1	0.2%
<b>Total</b>	<b>129</b>	<b>100.0%</b>	<b>177</b>	<b>100.2%</b>	<b>127</b>	<b>99.9%</b>	<b>304</b>	<b>100.0%</b>	<b>433</b>	<b>100.0%</b>

percent of the total citations, 29.5 percent of the citations captured by ISI, and 51.3 percent of the print citations, but only 2.4 percent of the Web citations. It is noteworthy, but not surprising, that the Web seldom cites documents published earlier than 1990 because it did not exist at that time.

The distribution of citations among cited documents was highly skewed, with 17.8 percent of the documents (8 of 45) accounting for 58.7 percent of the citations (254 of 433), whereas 28.9 percent of the documents (13) produced 74.4 percent of the citations (322). This pattern roughly approximates, but does not fully conform to, Richard L. Trueswell's well-known 80/20 Rule, modeled on the economist Vilfredo Pareto's research, whereby 20 percent of a library's books account for 80 percent of the circulation.<sup>32</sup> It is interesting—and possibly significant—that ISI captured a smaller proportion of the total citations to highly cited documents than it did for all the other lesser-cited documents. Specifically, only 27.6 percent of the 254 citations to the eight most frequently cited documents were captured by ISI, whereas it captured 33.0 percent of the citations to all the remaining documents. The complete explanation for this phenomenon is unclear. However, it reflects a pattern reminiscent of Bradford's Law, in which citations to highly cited documents are scattered among many citing documents including a relatively larger proportion of marginal ones beyond ISI's reach, whereas citations to lesser-cited documents are concentrated in a smaller number of core-type publications more likely to be covered by ISI.<sup>33</sup>

### Summary and Conclusions

Because the author is familiar with citations to his work and has meticulously recorded them for years, this project offers the benefit, to borrow wording from Cronin and Shaw, of "thick description" by an "insider" with "intimate knowledge."<sup>34</sup> This investigation found that ISI's coverage of the author's total citations is

higher than the 16 percent rate reported by Van Hooydonk and Milis-Proost for Ghent professors, but the coverage of journal citations is lower than the 74 percent reported by Funkhouser.<sup>35,36</sup>

It is assumed that each author has his or her own citation profile and that this case study offers one piece of a broader mosaic concerning the parameters of ISI coverage of author citedness. Questions requiring further research include:

- How does one define a Web citation?
- What is the best method for classification of Web citations?
  - How effective are the various Web search engines in retrieving Web citations?
  - How would rankings of top programs or most productive scholars differ if non-ISI print citations were included?
  - How would the rankings differ if Web citations were included?

This research project's salient findings and tentative conclusions are outlined below:

- The ISI databases, with self-citations disregarded, captured approximately 40 percent of the author's print citations and about 30 percent of his total citation count (Web citations included).
  - Approximately half of the book reviews of the author's work were included in the ISI databases, but instances in which his contributed chapters to edited books were mentioned in book reviews could not be retrieved by searching the author's name.
    - ISI included citations in only two of the fifteen languages and from six of the twenty-eight countries represented in the author's total citation portfolio.
    - Any LIS author using citation data in a promotion and tenure dossier or wishing to identify his or her full citation record for any purpose would be well advised not to rely exclusively on the Web of Science.
      - Rankings based on ISI data of a discipline's most-cited authors or academic departments might be significantly different if non-ISI citation data were included.
      - Web citations display a substantially different profile from those captured by ISI and print citations in general be-

cause the Web includes a broader range of languages and countries of origin, is more likely to cite unrefereed journals articles, and seldom cites documents published before 1990.

Inevitably, one cannot be certain that the author identified all non-ISI citations. Therefore, if this study's data err, they err in overestimating rather than underestimating the proportion of citations covered by ISI.

This case study illustrates the ongoing transformation of scholarly communication via technology, a topic of interest to the Association of Research Libraries and other organizations. A decade ago, citations were almost inevitably print, but more than 29 percent of this study's citations were from the Web. The finding that 29.1 percent of the Web citations were to nonrefereed journal articles (11.4% of the print citations were to nonrefereed articles) suggests that nonrefereed publications are playing a greater role in scholarly communication on the Web than in print and is consistent with the conventional view regarding the Web's lack of quality control. That nearly 30 percent of the Web citations originated outside the United States, with twenty-four countries represented, suggests a globalization of scholarly communication on the Web. Ironically, these are contradictory trends

from the promotion and tenure perspective, which often discounts unrefereed publications but highly values international impact. Promotion and tenure committees have wrestled with the question of whether a publication in an exclusively electronic format counts as much as one in print. Similar issues might arise regarding print and electronic citations, but the eventual full-scholarly acceptance of both electronic publications and citations (after a consensus on their definition is reached) seems highly probable.

Finally, this investigation has significant implications for academic librarians who use the Web of Science to help compile their personal promotion and tenure dossier, for faculty members who need assistance in compiling their own citation records for whatever reason, and for use as a general reference tool. Although it is generally recognized that ISI authorship citation data are incomplete, this project examines the parameters of ISI coverage and demonstrates that it is especially inadequate in identifying citations from non-English sources and from sources originating outside the United States (with the exception of the United Kingdom and the Netherlands). Thus, the ISI should never be relied upon exclusively when these types of citations are important.

---

### Notes

1. Kathlyn L. Reed, "Citation Analysis of Faculty Publications: Beyond *Science Citation Index* and *Social Science* [sic] *Citation Index*," *Bulletin of the Medical Library Association* 83 (Oct. 1995): 507.
2. Edward Truman Funkhouser, "The Evaluative Use of Citation Analysis for Communications Journals," *Human Communications Research* 22 (June 1996): 563-74.
3. Barbara A. Rice and Tony Stankus, "Publication Quality Indicators for Tenure or Promotion Decisions: What Can the Librarian Ethically Report?" *College & Research Libraries* 44 (Mar. 1983): 173-78.
4. Reed, "Citation Analysis of Faculty Publications," 503-8.
5. Blaise Cronin and Kara Overfelt, "Citation-based Auditing of Academic Performance," *Journal of the American Society for Information Science* 45 (Mar. 1994): 61-72.
6. Pamela S. Bradigan and Carol A. Mularski, "Evaluation of Academic Librarians' Publications for Tenure and Initial Promotion," *Journal of Academic Librarianship* 22 (Sept. 1996): 360-65.
7. The advantages and disadvantages of citation data in general, most of which are applicable to authorship data, are discussed in Thomas E. Nisonger, "The Benefits and Drawbacks of Impact Factor for Journal Collection Management in Libraries," forthcoming in *Serials Librarian*.
8. The author has served for five years as chair of his school's Promotion and Tenure Committee.
9. Howard D. White, "Toward Ego-centered Citation Analysis," in *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, ed. Blaise Cronin and Helen Borsky Atkins (Medford, N.J.: Information Today, 2000), 475-96.
10. Blaise Cronin and Debora Shaw, "Identity-creators and Image-makers: Using Citation

Analysis and Thick Description to Put Authors in their Place," in *Proceedings of the 8th International Conference on Scientometrics & Informetrics: Sydney, Australia, July 16–20th 2001*, vol. 1, ed. Mari Davis and Concepción S. Wilson (Sydney Australia: The Bibliometrics and Informetrics Research Group, the University of New South Wales, 2001), 136.

11. Christine L. Borgman and Jonathan Furner, "Scholarly Communication and Bibliometrics," in *Annual Review of Information Science and Technology*, vol. 36, ed. Blaise Cronin (Medford, N.J.: Information Today, 2002), 3–72.

12. Bernie Sloan, "Personal Citation Index: Exploring the Impact of Selected Papers." Available online from <http://alexia.lis.uiuc.edu/~b-sloan/pci2.html>.

13. Heting Chu, "Intellectual Activities and Influences of Belver C. Griffith: A Citation Perspective," *Scientometrics* 51 (Mar.–Apr. 2001): 481–88.

14. Michael J. Moravcsik, "Citation Context Classification of a Citation Classic Concerning Citation Context Classification," *Social Studies of Science* 18 (Aug. 1988): 515–21.

15. Michael J. Moravcsik and Poovanalingam Murugesan, "Some Results on the Function and Quality of Citations," *Social Studies of Science* 5 (Feb. 1975): 86–92.

16. Robert M. Hayes, "Citation Statistics as a Measure of Faculty Research Productivity," *Journal of Education for Librarianship* 23 (winter 1983): 151–72.

17. John M. Budd and Charles A. Seavey, "Productivity of U.S. Library and Information Science Faculty: The Hayes Study Revisited," *Library Quarterly* 66 (Jan. 1996): 1–20.

18. John M. Budd, "Scholarly Productivity of U.S. LIS Faculty: An Update," *Library Quarterly* 70 (Apr. 2000): 230–45.

19. Deborah Lynne Wiley, "Cited References on the Web: A Review of ISI's *Web of Science*," *Searcher* 6 (Jan. 1998): 32–39.

20. Victor Rosenberg, "An Assessment of ISI's New *Web of Science*," *Information Today* 15 (Mar. 1998): 21, 61.

21. Barbara W. Quint, "New ISI *Web of Knowledge* Integrates *Web of Science* with Current Contents Connect," *Information Today* 18 (July/Aug. 2001): 42.

22. Funkhouser, "The Evaluative Use of Citation Analysis," 565–67.

23. G. Van Hooydonk and Greta Milis-Proost, "Measuring Impact by a Full Option Method and the Notion of Bibliometric Spectra," *Scientometrics*, 41 (Jan.–Feb. 1998): 169–83.

24. Pam Royle and Ray Over, "The Use of Bibliometric Indicators to Measure the Research Productivity of Australian Academics," *Australian Academic & Research Libraries* 25 (June 1994): 77–88.

25. Selden Durgom, "A Comparison of Faculty Publications to Publication Records in ISI Citation Indices: How Well Does ISI Reflect the Publications of LIS Faculty?" (Master's thesis, University of North Carolina at Chapel Hill, 1998).

26. The *Web of Science*, unlike the earlier ISI indexes, is supposed to include citations to an author's work regardless of whether he or she is the first-named or a subsequent author. Although anecdotal evidence reports that subsequently named coauthors are not fully covered in the *Web of Science*, this investigator found that all citations to articles in which he was the second author were retrieved under his name (as well as the first author's name).

27. Thomas E. Nisonger, *Collection Evaluation in Academic Libraries: A Literature Guide and Annotated Bibliography* (Englewood, Colo.: Libraries Unlimited, 1992).

28. ———, *Evaluation of Library Collections, Access, and Electronic Resources: A Literature Guide and Annotated Bibliography* (Westport, Conn. and London: Libraries Unlimited, Greenwood, 2003).

29. Reed, "Citation Analysis of Faculty Publications," 506.

30. Blaise Cronin et al., "Invoked on the Web," *Journal of the American Society for Information Science* 49 (Dec. 1998): 1319–28.

31. Liwen Vaughan and Debora Shaw, "Bibliographic and Web Citations: What Is the Difference?" *Journal of the American Society for Information Science and Technology* 54 (Dec. 2003), 1313–32.

32. Richard L. Trueswell, "Some Behavioral Patterns of Library Users: The 80/20 Rule," *Wilson Library Bulletin* 43 (Jan. 1969): 458–61.

33. S. C. Bradford, "Sources of Information on Specific Subjects," *Journal of Information Science* 10, no. 4 (1985): 178. Technically, this pattern is a reversal of Bradford's Law, which posits concentration in highly cited journals and scatter among infrequently cited titles.

34. Cronin and Shaw, "Identity-creators and Image-makers," 33.

35. Van Hooydonk and Milis-Proost, "Measuring Impact by a Full Option Method," 169.

36. Funkhouser, "The Evaluative Use of Citation Analysis," 567.