Practicing Reference . . .

My Year of Citation Studies, Part 1

Mary Whisner

Ms. Whisner begins a year of exploring how legal scholarship citation counts are created and viewed. What works do authors actually cite? Which legal sources are included? She shares her first findings here.

Introduction

¶1 Citation runs through the law and librarianship. Much of reference and research is helping people find sources to cite. We also help our professors find sources that cite them.1 And we help people with citation format.2 Scholars have studied citation patterns from many angles,3 sometimes using citation counts to form judgments about the quality or influence of journals,4 articles,5 authors,6 or faculties.7

* © Mary Whisner, 2018. I am grateful to Ashley Arrington, Gerard Fowkes III, and Matthew Neely for slogging through footnote-heavy articles to extract citations. My thanks also to Todd Weldermuth for asking questions that inspired some of this work. For commenting on a draft, I thank Crystal Alberthal, Anna Endter, Maya Swanes, and Nancy Unger.


1. For tips, see Mary Whisner, Tracking Citations to Articles, Gallgher Law Libr., Univ. of Washington Sch. of Law, http://guides.lib.uw.edu/law/citestoarticles (last updated July 5, 2017).


3. In 1999, the West Group sponsored “Interpreting Legal Citations,” a symposium held at Northwestern University. The papers, published in the Journal of Legal Studies, 29 J. Legal Stud. 317–584 (2000), look at a variety of questions using different methods. (I have that issue of the Journal of Legal Studies in print, so I see the introduction and eleven papers as a collection. Readers today would likely download one or more of the articles from this issue along with various articles from other journals, and so would not have that perception.)


7. See, e.g., Benjamin H. Barton, Is There a Correlation Between Law Professor Publication
I can’t pretend to have read all of the bibliometric studies of legal scholarship, although I think they’re fascinating. They address many interesting questions: Who (if anyone) is using the articles that our faculties and student authors are churning out? What are the most successful articles? How long does an article stay in the limelight?

But I’m also aware of the limitations of citation studies. What is counted depends on the databases available and the way they are searched. Moreover, frequent citation is an imperfect proxy for usefulness, brilliance, or importance. For example, suppose author A writes a great article about distribution of assets in one of the nine community property states when a nonmarital relationship breaks up. The article could be useful to practitioners in those states and interesting to family law scholars—but, despite its qualities, that article won’t reach the stratosphere of citation counts because it’s a narrow topic, made narrower by its application in only nine states. Now consider author B’s article about constitutional interpretation. It touches on many hot-button topics (abortion, free speech, gun control, presidential power), but it’s a little sloppy. Many of the people who cite it disagree with it. Later citations cite the critiques and add parentheticals that the article they’re citing quotes author B. Author B’s paper may have twice as many citations as author A’s, but we shouldn’t conclude that it is better.

Sometimes I come up with questions that might be addressed by a citation study. How many professional articles cite student notes and comments? How many student works cite student works? Is anybody citing legal encyclopedias? Has the widespread use of journal articles in electronic format changed citing preferences? (That is, if it’s just as easy to find and download an article from a nonelite school’s journal as from the *Yale Law Journal*, has the mix of cited journals shifted at all?) How often are law review articles cited in briefs later cited in a court opinion? And how often do courts find and cite articles on their own without the briefs having cited them first?

I have questions, but I don’t have the knowledge, time, or skills to do big, complex studies. For example, Ian Ayres and Fredrick E. Vars created a dataset of 979 articles (excluding student pieces) from three journals over sixteen years, and


People who create citation-based studies are aware of this. They don’t claim that citation count is a perfect measure of scholarly quality, just that it is an objective measure that can be used. See, e.g., Sisk et al., Scholarly Impact 2015, supra note 7, at 116 (making the point that to be useful, citation counts only need to correlate with quality, not precisely parallel it).
coded the articles by subject, position in the issue, and other characteristics. They examined citation counts (from Social Science Citation Index) for the articles, using regressions to control for years after publication, journal, and subject. The result\(^9\) is impressive and interesting. And, because of its scope and sophisticated statistical analysis, it’s completely out of my league.

\(\%6\) Sometimes, though, I can poke around in a small sample and observe something interesting—and sometimes even useful. Last June, the faculty member teaching a writing seminar for our school’s environmental law journal\(^10\) asked me to find the most-cited student works on environmental law. He thought that this would be a good step in discussing with the student what makes a successful note or comment. I did some digging and gave him lists of the top pieces since 2011 and the top pieces of all time. He found it useful, and one of the students in the seminar has twice told me that he thinks it’s really valuable. I will talk more about my methodology later, but for now I’ll just assure you that there were no chi-squares, coefficients, or standardized residuals to be found. I later did a similar project when he taught a seminar for a different specialty journal.

**The Year of Citation Watching**

\(\%7\) A thorough, broad (and yet detail-rich) citation study is more than I can tackle now, even if I knew much more about quantitative research methods than I do. But I think the “Practicing Reference” is a good forum for sharing some small-scale bibliometric explorations. This installment is a start. There will be more, though: I have enough questions—and even tentative findings—to keep going. And so I declare 2018 to be a year of citation watching, at least for purposes of this column.\(^11\)

**The Requested Note and Comment Studies**

\(\%8\) Working on the professor’s request for the most-cited environmental law student works, I encountered several methodological challenges. First, notes and comments don’t always proclaim themselves in their titles. I like the ability in HeinOnline to sort by number of times cited by articles, but I found that HeinOnline’s labels—[article], [note], [comment]—were not reliable. Sometimes things labeled as notes or comments were commentary by law professors, speeches by judges, or other works that definitely were not by students. The author or journal may even label a piece “comment,” even though it is not a traditional student comment.\(^12\) On the other hand, some works HeinOnline labeled as articles were by students (skimming the first couple of pages led to the phrase “this Note” or “this Comment”).\(^13\) Second, I knew

---

11. The “Practicing Reference” column first appeared in volume 91, number 2, so this is the column’s twentieth year. Speaking of counting.
13. When I came across these mislabeled works, I used HeinOnline’s feedback form to tell them about what I saw as an error. Hein has always been responsive to feedback, and I had an e-mail correspondence and an in-person chat with Shane Marmion, now president of William S. Hein & Co., Inc.
that not all environmental law articles would use the word “environment” in the title. At the same time, some titles with the word “environment” are talking about the “environment of business” or the “regulatory environment.”

¶9 I searched for environment* OR natural resource* OR pollut* OR conservation in the title. I sorted by “Number of Times Cited by Articles,” and I skimmed. I didn’t restrict the search to Notes and Comments and, in fact, did find some things labeled articles that were notes. I knew that my search was not comprehensive. For example, it would miss notes with “Endangered Species Act” or “Superfund” in the title, if the titles did not also use one of my search terms.

¶10 I did not think to use HeinOnline’s Subject field. Using the subject Environment/Conservation Law (in addition to searching for a common word, like “law,” in the text) would have obviated the need to guess at “environmental-ish” words in the title. But this tool is also both underinclusive and overinclusive. For instance, it misses the most-cited note I found last year.¹⁴ And yet it includes an article about client-centered counseling¹⁵ whose text happens to use the word “environment” a lot.¹⁶ If I were to do this search again, I might do two searches: words in the title and subject.

¶11 I gave the professor a list of the twenty-two most-cited student works. I was frank about why I chose twenty-two: I stopped adding to the table when my shift ended. I made a second list with the eleven student works published since 2011 that had been cited at least four times. In the interest of space, just the top five from each list are in tables 1 and 2 here. I note the year of publication to make it easy to see that some of the most-cited works have been around for decades. It’s harder to pile up citations in just a few years.¹⁷ Four out of five of the most-cited works since 2011 were published in 2011. Surely in a few years some of the pieces from more recent years will catch up.

¶12 The same instructor was recruited to teach the writing seminar for the school’s online-only technology journal¹⁸ and asked me for a similar list. This time, I looked for notes and comments on technology and the arts from any journal.¹⁹ I limited the search to works HeinOnline had tagged as notes or comments. I knew that the tags were often inaccurate, but this time I chose not to invest the time it would take to skim author affiliations.

¶13 I also created a separate list for student pieces from online technology journals. I started with a list of online-only law journals and skimmed for relevant topics. That gave me seventeen journals, fifteen of which were available on HeinOn-

---


¹⁶. E.g., id. at 349 (“law school environment”); id. at 352 (“legal services environment”); id. at 358 (“predominantly black environment”); id. at 366 (“cultural environment”).

¹⁷. See Ayres & Vars, supra note 9, at 430 (“The number of citations an article receives is obviously related to how many chances it has had to be cited.”). Ayres and Vars used a regression analysis to make up for this bias in older articles on citation studies. I just invite you to eyeball the lists.


¹⁹. My search was art OR music OR tech* OR computer* OR digital* OR cyber* OR patent* OR robot* OR drone* OR automation in the title.
I searched for the word “law” in those fifteen journals, sorted the list by times cited, and skimmed. I noticed “articles” written by students from schools other than the journal’s home school. I decided to include these in the list because the point of the project was to find successful student writing. The top entries from the two tech lists are in tables 3 and 4.


21. I did not check to see whether a given journal was online-only at the time of a piece’s publication.

<table>
<thead>
<tr>
<th>Year</th>
<th>Student Work</th>
<th>Times Cited in Journals</th>
<th>Times Cited in Cases</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Student Work</th>
<th>Times Cited in Journals</th>
<th>Times Cited in Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Emily Sangi, Note, <em>The Gap-Filling Role of Nuisance in Interstate Air Pollution</em>, 38 Ecology L.Q. 479 (2011)</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>
¶14 It was only when I was creating table 4 that I noticed the coincidence that three pieces were from the same volume of the *Columbia Science and Technology Law Review*, and they each had the same number of citations. Closer inspection revealed that all of the pieces in that volume—by professional authors and by students—had the same citation: 4 COLUM. SCI. & TECH. L. REV. 1. I began to suspect that all of them together had thirty-four citations, not that each of them did. So instead of relying on HeinOnline’s ScholarCheck tally, I ran searches for the author’s name and words in the title (table 5). My guess was half right: not one of the works had thirty-four citations. But I was surprised to see that the total was forty-four, not thirty-four. Surely there wasn’t that much overlap in citing references. Plus there would be even more citations if we looked at the outside articles in that volume (that had the same citation).

¶15 I dug a little deeper and found that many of the citing references weren’t in Bluebook form. For example, John Miller’s piece is cited as:


You get the point. If HeinOnline’s algorithm looks for the Bluebook citation, then that would explain both the original miscount (attributing the thirty-four citing references to each piece with the same citation) and the undercount.

This potential for undercounting applies to all citation counts in the system. Let’s go back to the most-cited student work in environmental law: Rachel D. Godsil, Note, *Remedying Environmental Racism*, 90 Mich. L. Rev. 394 (1991). Last June,
HeinOnline’ ScholarCheck told me it was cited by 157 articles and two cases. In November 2017, it was 158 articles. When I click on the link for “Cited by 158 Articles,” HeinOnline displays the search it uses to generate the list of citing articles: 158 results searching for ("90 Mich. L. Rev. 394" OR "90 Mi. L. 394" OR "90 Mich LR 394" OR "90 Michigan Law Review 394") AND NOT id:hein.journals/mlr90.21) in Law Journal Library.


§19 These fifteen citing references were missed by HeinOnline’s algorithm because the journals used different citation styles (as with the nanotechnology example), because HeinOnline’s OCR of the original was garbled, or—in one instance—because the citing author got a page wrong.

**Variant Citation Form Examples**


---


### Jumbled OCR Examples

- 90 MicH. L. R-v. 394\(^{33}\)
- 90 MICH. L. Rnv. 394,400\(^{34}\)

### Typo in Citing Reference


\(\S 20\) I draw several lessons from this project. First, it is possible to use tools we have to develop lists of frequently cited works. Second, it’s wise to look beneath the surface to check the results, as when I noticed that some articles were tagged as notes or comments and vice versa. Third, researchers should be aware of ways that algorithms can provide results that are inaccurate. Being aware of the inaccuracies should not prevent us from using the tools we have: they are great helps, even if with their inaccuracies. And, for most projects, we shouldn’t feel compelled to do multiple searches to scour away the inaccuracies. Precision has costs. Double-checking each citation in the lists I gave the professor would have taken ages, diverting me from my other work (not to mention making me either nuts or resentful). And the product would not have been more useful to the professor and his students.

### What Do Authors Cite?

\(\S 21\) I am curious about what authors cite. What is the balance of law review articles, notes and comments, books, cases, statutes? To explore this, I thought I’d take a small sample of works—say, five articles and five student pieces—and look. I created two lists, the twenty most-cited articles (table 6) and the twenty most-cited

---


<table>
<thead>
<tr>
<th>Rank</th>
<th>Article Identifier</th>
<th>Citation</th>
<th>Cited by Articles (HeinOnline)</th>
<th>Other Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 (tie)</td>
<td>13</td>
<td>Rainier Elias, <em>Identity, Law, and Essentialism: What’s Love Got to Do with Same-Sex Marriage?</em>, 6 Crit 1 (2012)</td>
<td>85</td>
<td>WS: n/a, SH: n/a, KC: 0</td>
</tr>
</tbody>
</table>
notes and comments from 2012, based on HeinOnline’s ranking. Why 2012? I figured that five years was a good length of time to get cited but wasn’t so far back in time that I wouldn’t see any citations to YouTube or Twitter.

¶22 In many fields, it would be easy to see all works an article cites because it’s standard to have a list of references at the end. But legal scholars put all their references in footnotes (a system I’m generally very comfortable with—except when I want a simple list!).

¶23 What I’d like is a table of authorities, like at the beginning of a brief, with the different authorities split out by type—e.g., Cases, Statutes and Regulations, Secondary Sources. Both KeyCite (in Westlaw) and Shepard’s (in Lexis Advance) enable researchers to see a table of authorities for a given work. Alas, they generally include only cases, and I’m interested in the entire range of authorities cited. Web

<table>
<thead>
<tr>
<th>Rank</th>
<th>Article Identifier</th>
<th>Citation</th>
<th>Cited by Articles (HeinOnline)</th>
<th>Other Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SH: 86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KC: 91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SH: 77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KC: 83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SH: 78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KC: 92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SH: 90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KC: 92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SH: 79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KC: 86</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>Annemarie Bridy, <em>Coding Creativity: Copyright and the Artificially Intelligent Author</em>, 2012 Stan. Tech. L. Rev. 5</td>
<td>80</td>
<td>WS: n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SH: 16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>KC: 13</td>
</tr>
</tbody>
</table>

36. In the sample of elite journals studied by Ayres and Vars, “[c]itations to a piece peaked 4 years after its publication, declined, then flattened out.” Ayres & Vars, supra note 9, at 436. I read that after I’d chosen 2012, but it’s nice to have my hunch validated.


38. Shepard’s and KeyCite include some IRS materials in the Table of Authorities for law review articles. For example, the Shepard’s and KeyCite tables of authorities (both accessed Nov. 13, 2017) for Note, *Taxing Private Equity Carried Interest Using an Incentive Stock Option Analogy*, 121 Harv. L. Rev. 846 (2008), each list seven authorities: three Revenue Procedures, two Revenue Rulings, and two Notices. The Note has ninety-four footnotes and does not cite a single case. Now I’m curious about how many law review pieces cite no cases. Perhaps that’s a rabbit hole to go down another time.
of Science enables one to generate a list of cited references in a work, but it doesn’t include statutes or cases.

¶24 I decided to tackle the problem with brute force. I downloaded the top article in the list from Westlaw.39 I love HeinOnline’s PDFs, but I wanted a Word version. When I’m reading an article, I find it annoying that Westlaw puts all the footnotes at the end, but that positioning was perfect for this task because I could copy and paste just the footnotes into another document. From there, I could start creating a table of authorities by harvesting citations from the footnotes and sorting them into groups. I disregarded all the *id.* and *supra* references because I wanted only one citation for each source. I compared my harvest with the lists generated by Shepard’s and Web of Science and decided it was worthwhile to use the labor-intensive but more thorough method.40

¶25 I enlisted the help of three of our law librarianship students. They found that the project was as tedious as I’d said when I recruited them,41 but they hung in there and generated tables of authorities. An outlier within this small sample of articles was the one that weighed in at 150 pages with 722 footnotes.42 Harvesting the authorities from that giant was more grueling than either the intern or I expected.

¶26 It will take me some time to sort through the tables of authorities the interns created, but even a surface look is enough to notice quite a range in citation choices. For example, an article on implicit bias cites more than three times as many articles as does an article on a civil procedure topic, and more than half are from journals outside law.43 This is not terribly surprising—you’d hope that people writing about a social science topic would use social science scholarship—but I think there’s some value in being able to pin a number on it.

Shepard’s does not appear to include administrative decisions in tables of authorities. For example, Michael D. Moberly, *Striking a Happy Medium: The Conversion of Unfair Labor Practice Strikes to Economic Strikes,* 22 Berkeley J. Emp. & Lab. L. 131 (2001), cites many NLRB decisions, but they are not in the Table of Authorities (accessed Nov. 13, 2017). KeyCite does list the NLRB decisions in its Table of Authorities. The “Type” column labeled authorities either “Case” or “Administrative Decision & Guidance.”

Another article, Pooja Shethji, Note, *Credit Checks Under Title VII: Learning from the Criminal Background Check Context,* 91 N.Y.U. L. Rev. 989 (2016), cites three EEOC decisions, *id.* at 996 n.32, 998 n.46, 1005 n.78. One shows up in the Table of Authorities (accessed Nov. 13, 2017), but the reader can’t tell it’s an EEOC decision. The author’s citation to “EEOC Decision No. 72-427, 4 Fair Empl. Prac. Cas. (BNA) 304, 1971 WL 3943,” *id.* 989 n.46, appears in the Shepard’s Table of Authorities under “Other Federal Decisions” as “4 Fair Empl. Prac. Cas. (BNA) 304.” KeyCite’s Table of Authorities (accessed Nov. 13, 2017) lists all three, with full citations.


40. One oddity: I found two cited cases that Shepard’s missed. Shepard’s found one case that I missed. But, upon examination, it turned out that Shepard’s listed a case, *San Luis Obispo Mothers for Peace v. Hendrie,* 11 Envtl. L. Rep. 20455 (D.D.C. 1980), that the article hadn’t cited at all.

41. It’s a wonderful thing when I send out an e-mail message asking for one or two interns to work on a tedious project and get four volunteers!


I also can see that there are challenges in deciding how to count some types of authority. It’s easy to count the number of law journal articles cited, but how should I count statutes? Do citations to five sections of a statute or regulation count as five citations or one? How about subsections? Should I count a citation to the Freedom of Information Act (5 U.S.C. § 552) separately from a citation to the Administrative Procedure Act (codified in scattered sections of 5 U.S.C., including §§ 551–559)? The ambiguities might make it too hard to come up with a meaningful tally. Maybe I’ll just look at the blunt question whether an article cites statutes at all. Does the article cite a federal statute, yes or no? Does it cite a state statute, yes or no?

I created my lists of most cited articles and student works in HeinOnline, using its numbers for the times cited by articles. But when I looked at other online tools for counting citing references, I saw some sharp differences. For example, the top article in the list was cited 160 times according to HeinOnline, but only 100 times according to Web of Science. Shepard’s and KeyCite were close, with 168 and 176 citations, respectively. A graph showing the comparative ups and downs is in figure 1.

Figure 1

Comparison of Citing Reference Counts for HeinOnline, Web of Science, Shepard’s, and KeyCite

Articles are ranked by the number of citations listed in HeinOnline’s Scholarcheck. The solid line indicates citations listed in Web of Science. Shepard’s and KeyCite results are shown with dashed and dotted lines, respectively.

44. The HeinOnline search was done Nov. 3, 2017. Web of Science searches were done Nov. 8, 2017. I generally searched for words in the title, adding other fields if title words were common. Shepard’s searches (Nov. 8 and 10, 2017) were by each article’s citation. The number is for “Other Citing Sources” (i.e., citing sources other than decisions), excluding “Court Documents” to count just law review articles and treatises. KeyCite searches (Nov. 8, 2017) were by article citation. The number is for “Secondary Sources.”
¶29 Why would the systems vary so much? First, Web of Science just doesn’t include as many law journals as the other three systems, so it misses citations. It has no entries at all for the articles that were published in the Columbia Business Law Review, The Crit: A Critical Studies Journal (University of Idaho College of Law), the Oregon Law Review, or the Stanford Technology Law Review. KeyCite and Shepard’s often show more citing references than HeinOnline because they include treatises and encyclopedias. They both fell short on The Crit, though: KeyCite had an entry that showed zero citing references, while Shepard’s had no entry at all.

Running Out of Steam, Running Out of Time

¶30 Obviously there is more to be explored, even in the data I’ve already assembled. But I am running out of steam and I really ought to send along this piece to my long-suffering editor. So I will recap by saying that I’ve demonstrated some ways to gather interesting material from HeinOnline and other tools, even without sophisticated data analysis tools. I’ll note that it’s useful to be aware of the limits of the tools (e.g., the ways that HeinOnline’s algorithm can miss citations or that Web of Science does not include all law journals). And I’ll close for now with the promise that I will write more later. This is just the first issue of 2018.