

Another Response to Taylor's Comparison of KeyCite and Shepard's*

Dan Dabney**

Mr. Dabney, one of the designers of the KeyCite service at West Group, comments on two aspects of the article by William Taylor in the Spring 2000 issue of Law Library Journal and reports the results of a small ancillary experiment.

¶1 We congratulate William Taylor on his article comparing KeyCite and Shepard's.¹ Such painstaking empirical studies require a great deal of careful work, but they generate concrete facts that are not to be had from more speculative or impressionistic articles.

¶2 Unfortunately such studies are very laborious. In designing such a study, one must of necessity adopt simplifying assumptions and streamlined procedures that can expose the work to methodological objections. It takes a brave scholar to expose his or her work to this kind of scrutiny.

¶3 With the publication of the article, along with a short response article that appeared in the same issue of *Law Library Journal*,² perhaps it is now appropriate for West to offer a brief response to Mr. Taylor's article.³

¶4 We have two points: one brief observation on the currency portion of the study, and one more extended comment on Mr. Taylor's methodology in the accuracy portion of the study. In connection with the latter point, we report the result of a small ancillary experiment, and suggest an alternate interpretation of the data on the basis of that experiment.

* © Dan Dabney, 2000.

** Senior Director for Research and Development, Legal Product Development, West Group, Eagan, Minnesota.

1. William L. Taylor, Comparing KeyCite and Shepard's for Completeness, Currency, and Accuracy, 92 L. LIBR. J. 127, 2000 L. LIBR. J. 13.

2. Jane W. Morris, *A Response to Taylor's Comparison of Shepard's and KeyCite*, 92 L. LIBR. J. 143, 2000 L. LIBR. J. 14.

3. Mr. Taylor provided West Group with advanced copies of two drafts of the article, but asked no specific questions and did not invite our comments. Under the circumstances, we felt it inappropriate to comment on the substance of the article. West Group had no desire to make the experience even more painful by exposing the author to a barrage of commercially motivated criticism.

Currency

¶15 The brief observation concerns Mr. Taylor's conclusion that KeyCite lags behind its competitor in the speed with which indirect history references are added to the database.⁴ This was certainly true last fall when Mr. Taylor was performing his experiment. At that time, West Group's case law editorial department was experiencing an unusual backlog. Since that time, West has implemented new procedures that help maintain currency in such peak periods. When there is more than a two-day inventory of cases waiting for full editorial treatment, including headnotes and synopses, we make a separate preliminary history scan. In this scan, each citation in each case is reviewed to see if it requires a history tag. Thus, if Mr. Taylor's currency study were to be replicated today it would, we trust, show that KeyCite's currency has been improved.

Accuracy

¶16 The methodological point has to do with the procedure Mr. Taylor uses to assess the relative accuracy of the history assignments for the two systems.

¶17 Mr. Taylor identifies a certain number of citation pairs as omissions of negative history because they appear in one system but not in the other.⁵ This procedure contains the lurking assumption that history errors are likely to be errors of omission rather than of commission—that one system or the other might well overlook a citation that should have a history tag, but that it is unlikely that any of the tags actually applied are in error.

¶18 If this assumption is not true it might alter the results of the experiment substantially. An incorrectly applied tag is not only *not* counted as an error for the system that included it, but is also counted as an error for the other system that (correctly) omitted it. If one system contained materially more incorrect negative references than the other, conclusions based on Mr. Taylor's method could be misleading.

¶19 Mr. Taylor is aware of the problem:

It is possible that some of these negative analyses are unique to one system because they are mistakes. Looking at the text of each citing opinion, I did find some that I thought were incorrectly identified as negative, but I have decided not to interpose my own judgment in this very subjective area.⁶

¶10 While we sympathize with Mr. Taylor's reluctance to make quality judgments about the history postings in his sample, we feel that it is not possible to make a fair assessment of citation accuracy without doing so.

4. Taylor, *supra* note 1, ¶¶ 18–20.

5. *Id.* ¶ 22.

6. *Id.* ¶ 22 n.17.

Table 1
Unique History Tags, Correct and Incorrect, by System

	Total unique history tags	Tags found correct	Tags found incorrect	Overall accuracy
KeyCite	32	20	12	35
Shepard's	36	21	15	33

¶11 Fortunately, by providing readers with lists of the citation pairs that were counted as errors,⁷ Mr. Taylor gives anyone who chooses the opportunity to make the necessary correction. We have performed such an evaluation.

Method

¶12 West Group commissioned five attorneys from the Legal Research Center, Inc., a Minneapolis-based legal research company that has performed legal research tasks for clients all across the country, including both West Group and LEXIS-NEXIS.⁸

¶13 Each attorney was given all 68 of the citation pairs that were identified in Mr. Taylor's study as negative history by either KeyCite or Shepard's but not by both. The attorneys did not know which of the services had identified the citation pair as having history and which had not.

¶14 Each of the five attorneys considered each of the citation pairs, making a judgment about whether a negative tag was called for. The votes of the panel were tallied, and the propriety of the questioned KeyCite and Shepard's tags was assessed according to whether the tags accorded with the judgment of the majority of the five.

Results

¶15 Table 1 summarizes the findings of the study. Mr. Taylor, by not evaluating the propriety of the history tags, found Shepard's slightly more accurate because it had 36 unique history tags to 32 for KeyCite. Of those 36 tags unique to Shepard's, however, 15 were found to be incorrect, and ought to have been counted as Shepard's errors rather than KeyCite errors. Of the 32 tags unique to KeyCite, 12 were incorrect and should have been counted as KeyCite errors rather

7. *Id.* app. A (Negative Analyses Identified Only by KeyCite), app. B (Negative Analyses Identified Only by Shepard's).

8. *Editor's Note:* For more information about the Legal Research Center, its staff, and the nature and scope of its business, see *Legal Research Center* (visited July 6, 2000) <<http://www.lrci.com/home.htm>>.

Table 2
Level of Agreement Among History Judges

Vote	Number of cases
5-0	32
4-1	18
3-2	18

than Shepard's errors. Overall, according to the five attorneys from the Legal Research Center, KeyCite edged out Shepard's by a total of 35 correct postings to 33.

¶16 In this experiment, there was a great deal of variation among the judgments of the five attorneys. As shown in table 2, all five agreed on their history judgment less than half of the time.

¶17 It is unlikely that this level of disagreement would be found for history relationships in general. The sample here consists of citations on which the citation editors who create KeyCite and Shepard's disagreed with each other, and presumably ordinary citations tend to be less uncertain. As Mr. Taylor notes, however, citation history tags can be difficult and subjective, and it is only to be expected that any set of citation judgments is likely to be at variance with the perceptions of individual attorneys.

¶18 One way of demonstrating the subjectivity is to consider the results of the experiment if the history choices made by editors at KeyCite and Shepard's had been compared against the judgments of individual attorneys. As shown in table 3, the overall accuracy of the two systems as it would be perceived by individual attorneys varies widely.

Table 3
Overall Accuracy Ratings by Individual Judges

	Accuracy KeyCite / Shepard's
Judge 1	25 / 43
Judge 2	37 / 31
Judge 3	41 / 27
Judge 4	36 / 32
Judge 5	38 / 30

¶19 Perhaps the most salient point to take from this exercise is that one should not rely too much on the history judgments supplied by the editors for either system. Each system missed history tags that were correctly identified by the other, and each applied some tags that were properly omitted by the other.

¶20 But, while both systems could undoubtedly do somewhat better in applying history tags, it is not to be expected that they can always anticipate the judgments of individual legal researchers. If individual researchers do not agree among themselves, there is no way that any system can always agree with all of them. Thus, while history tags are of immense assistance to the legal researcher, they should not be relied upon exclusively. Careful researchers should also examine citations that do not have negative tags.

¶21 For much-cited cases this may be a forbidding task, but the researcher has many tools to help identify the cases that merit special attention. In KeyCite, the researcher can direct attention to the citations in the higher depth-of-treatment categories; cases that discuss the most relevant headnotes, the most recent cases, and cases from the most relevant jurisdictions.

Conclusion

¶22 The development of legal citators, like that of all of the other parts of the bibliographic apparatus of the law, owes much to the informed criticism of scholars like Mr. Taylor. We hope that he and others will continue to study these issues, and that when they do they will find that steady progress is being made.