



Measuring Legal Change: The Reliability and Validity of Shepard's Citations

James F. Spriggs, II; Thomas G. Hansford

Political Research Quarterly, Vol. 53, No. 2. (Jun., 2000), pp. 327-341.

Stable URL:

<http://links.jstor.org/sici?sici=1065-9129%28200006%2953%3A2%3C327%3AMLCTRA%3E2.0.CO%3B2-0>

Political Research Quarterly is currently published by University of Utah.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/utah.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.



Measuring Legal Change: The Reliability and Validity of *Shepard's Citations*

JAMES F. SPRIGGS, II, AND THOMAS G. HANSFORD,
UNIVERSITY OF CALIFORNIA, DAVIS

With few exceptions, scholars have generally relied on judges' final votes on the merits as the primary indicator of judicial outcomes. Yet, to fully understand judicial decision-making we think it imperative that research also focus on the interpretation of precedent and legal change. To do so, it is necessary to develop measures of legal change and the treatment of precedent over time. Scholars have begun doing so by using *Shepard's Citations*, a legal citation index. One of the most important features of *Shepard's* is its list of all opinions that legally treat a previously decided case, as well as its characterization of the nature of that legal treatment. Yet, the reliability and validity of *Shepard's* is unknown, and we should therefore be appropriately skeptical of it. This article empirically tests the reliability of *Shepard's* and discusses the validity of its coding protocols. Our analysis demonstrates that *Shepard's* coding of legal treatment is quite reliable, though there is some notable variance across *Shepard's* treatment categories. We also point out several features of *Shepard's* that could potentially affect the validity of a measure derived from it. We conclude that, as long as scholars keep these validity issues in mind, *Shepard's* can be a highly appropriate data source.

Since the 1940s, with the publication of C. Herman Pritchett's *The Roosevelt Court* (1948), scholars interested in courts have relied on judges' final votes on

NOTE: Spriggs recognizes funding from the U.C. Davis Institute of Governmental Affairs/IGCC Research Fellows Program for collection of the *Shepard's Citations* data. We also thank Jim Beach, Wes Duenow, Larry Eichele, Randy Gee, Robin Hastings, Danielle Perry, Brandon Reeves, David Richardson, Jamie Scheidegger, and Danny Williams for research assistance in gathering the *Shepard's Citations* data. We thank Saul Brenner and John Gates for helpful comments; and we further thank the Shepard's Company (especially John Strand and Jane Morris) for invaluable assistance in understanding *Shepard's* coding protocols and for helpful comments. Both authors contributed equally to this research.

Political Research Quarterly, Vol. 53, No. 2 (June 2000) pp. 327-341

the merits (i.e., whether supporting the liberal or conservative position) as the primary indicator of judicial outcomes. By examining individual judges' votes (e.g., Rohde and Spaeth 1976) or collective court outcomes (e.g., George and Epstein 1992), this approach has generated a considerable body of knowledge regarding the causal forces underlying case dispositions. Increasingly, however, scholars recognize that to understand judicial decision-making fully we must move beyond votes and study what is arguably the judiciary's most important policy output—the precedents set by court opinions.

This recognition has led scholars to develop ways of studying precedent from both an endogenous and exogenous perspective. From an endogenous point of view, scholars often attempt to explain why law changes over time (e.g., Wahlbeck 1998, 1997; Epstein and Kobyłka 1992) and thus consider changes in precedent as a dependent variable to be explained. Researchers also approach precedent exogenously, conceptualizing it as an independent variable likely to influence either court decisions (Knight and Epstein 1996) or decision-makers external to courts (Spriggs 1996). Indeed, one of the most important contemporary debates among judicial scholars is whether or to what extent the norm of *stare decisis* constrains judges by causing them to follow the legal rules established by prior decisions (e.g., Spaeth and Segal 1999; Gates and Phelps 1996; Knight and Epstein 1996; Brenner and Stier 1996; Songer and Lindquist 1996).

To examine empirically the treatment or effect of precedent, scholars must develop measures that tap how it changes. Yet, to measure changes in precedent as either dependent or independent variables appears a daunting task. The difficult and time-consuming nature of this undertaking partially explains why scholars generally rely on changes in judges' votes on the merits over time as a measure of legal change (e.g., Baum 1988; Segal 1985). While obviously indicative of the overall policy orientation of courts, such a measure does not truly capture changes in law.¹

Researchers have therefore begun utilizing *Shepard's Citations* to derive measures of legal change and the treatment of precedent (e.g., Spriggs and Hansford n.d.; Kemper 1998; Benesh and Reddick 1998). As discussed below, *Shepard's Citations* is a citation index that, among other things, provides a list of all U.S. court opinions that refer to any U.S. state or federal court case decided since the

¹ Besides judges' votes, past research offers two additional measures of legal change. First, Brenner and Spaeth (1995) identified the U.S. Supreme Court cases in which the most dramatic form of legal change, the overruling of a precedent, has occurred. While most likely both valid and reliable, this measure only taps one somewhat infrequent form of legal change. Second, Wahlbeck (1997, 1998) presents a highly innovative approach for measuring legal change, in which he examines alterations in the set of factual circumstances included or excluded in a legal rule. His method provides a reliable way to capture legal change, but it requires analysts to focus on relatively narrow issue areas. Social scientists therefore still need a technique that can allow for the study of changes in precedent across all issue areas in all cases decided in U.S. courts.

beginning of the U.S. legal system. *Shepard's* also indicates how a particular court opinion is legally interpreted by the subsequently decided cases that cite it.

Thus, one can use *Shepard's*, for instance, to ascertain how the Supreme Court has interpreted a particular precedent. To illustrate, *Shepard's* indicates that during the Warren Court no Supreme Court majority opinion (out of 8) dealing with *Miranda v. Arizona* (1966) interpreted the precedent in a negative manner, while during the Burger and Rehnquist Courts 37.9 percent (11 of 29) and 14.3 percent (1 of 7) respectively did so.² With this analysis, it is possible to measure both the extent to which the Court has altered a precedent and the vitality of a precedent at any particular time.

As social scientists, however, we must approach with skepticism any potential data source whose validity and reliability are unknown. In this context, validity refers to the extent to which *Shepard's* legal treatment categories actually correspond to the concepts they are intended to capture, while reliability concerns the reproducibility of the data (Carmines and Zeller 1979; Johnson 1987; Gates 1990). Indeed, the veracity of a study's results ultimately depends in part on the accuracy of the data. For this reason, judicial scholars have begun to evaluate closely the measures used in research (e.g., Epstein and Mershon 1996). If the data are unreliable or invalid then the study's empirical findings are highly questionable, regardless of the power of the theory or the sophistication of the research design. Yet, no study has empirically assessed the reliability or validity of *Shepard's* editorial analysis of court opinions. Therefore, we empirically test the reliability of *Shepard's Citation's* analysis of Supreme Court opinions and assess the validity of *Shepard's* treatment codes. Our results indicate that, with certain caveats that must not be ignored, *Shepard's* offers a valuable source from which to develop indicators of how Court opinions interpret previously decided cases. Researchers can therefore use *Shepard's* to derive measures of how precedents are interpreted in the U.S. over time.

THE MECHANICS OF SHEPARD'S CITATIONS

For each published state and federal court case decided since the beginning of the U.S. legal system (which we refer to as a "cited" case), *Shepard's* provides a list of all subsequently decided cases which refer to it (which we term "citing" cases). *Shepard's* also offers editorial analysis regarding how the citing case legally interprets the cited case.³ The question *Shepard's* asks in each case is the following: "What effect, if any, does the citing case have on the cited case?" (*Shepard's*

² *Shepard's* also reports that an additional 28, 68, and 19 majority or concurring Court opinions cited, but did not substantively legally treat, *Miranda* during the Warren, Burger, and Rehnquist years, respectively.

³ *Shepard's* provides the editorial analysis for a citing case's majority and concurring opinions, while for a dissenting opinion *Shepard's* only notes whether it refers to a previously decided case.

1993: 13). *Shepard's Citations* offers nine possible ways a citing case can legally treat a cited case: Overrule, Question, Limit, Criticize, Distinguish, Follow, Parallel, Explain, or Harmonize.

The Shepard's Company, which publishes *Shepard's Citations*, hires and trains attorneys to content analyze court opinions, a process it terms "letter editing." These legal editors receive extensive training based in part on the coding rules set forth in an in-house, unpublished training manual. This manual, which Shepard's graciously provided us, includes nearly 25 single-spaced pages of material on the overall "letter editing" process, with 13 pages strictly devoted to laying out the coding rules for each of the treatment categories (Shepard's 1993: 1-25). Thus, Shepard's recognizes the subjective nature of the coding enterprise and has developed coding rules to try and maximize data reliability.

Shepard's training manual's coding protocols consist of definitions and examples for each treatment category. To familiarize the reader with them, we will provide brief examples for each treatment type. Before moving to the nine treatment categories, we should note that *Shepard's* contains an implicit tenth treatment category: a citing case can refer to a cited case without legally interpreting it. If *Shepard's* lists a citing case as referring to a cited case but no treatment code is provided then this means the citing case contained a reference to, but did not legally treat, the cited case. In other words, it represents a non-substantive treatment of a cited case (Johnson 1986).

We begin with the two categories *Shepard's* labels "positive" treatment, meaning the citing case relied on the cited case as legal authority. First, *Shepard's* uses Followed to indicate that a citing case's majority opinion "expressly" relied on a cited case as precedent.⁴ According to the manual, Followed is only to be applied if the citing opinion contains language that goes beyond a "mere going-along" with the cited case. Examples of language requiring a Followed treatment are: "controlling," . . . "determinative," . . . such a conclusion is required by. . ." (Shepard's 1993: 17). Second, *Shepard's* uses Parallel when the citing case indicates that the cited case is "identical, on all fours, or parallel to" the cited case (Shepard's 1993: 21). The manual further states that Parallel is not assigned if the citing case just states a similarity between it and the cited case.⁵

With regard to the negative treatment of precedent, *Shepard's* includes four codes that it considers "strong" negative treatment of a cited case, or situations in which the cited case may have been negatively affected by the citing case

⁴ A concurring opinion cannot be coded as Following a cited case (Shepard's 1993: 17). If a concurrence uses a cited case as precedent then *Shepard's* would code it as having Explained the cited case (phone conversations with Leslie Martin (6-8-98) and John Strand (8-6-98), Shepard's employees).

⁵ According to John Strand (phone conversation 11-10-98), a Shepard's employee, Parallel is no longer used as a treatment category, and instead such cases are coded as Followed.

(Shepard's 1993: 14, 24). Overruled is the strongest form of negative legal treatment. *Shepard's* employs this code when the citing case explicitly shows intent to overrule the cited case. To apply this code, terms such as "Overrule" or "Disapprove" or other "expressed and specific language that clearly indicates the cited case has been Overruled" must be contained in the citing opinion (Shepard's 1993: 20).⁶ Second, *Shepard's* uses Questioned when the citing case determines that the cited case lacks continuing validity and thus can no longer be relied upon. It is appropriate, for example, where the citing case notes that the cited case has been previously overruled (either by a court or statute), and the citing case agrees with the overruling, or discusses two opposing lines of cases and aligns itself with one and notes that the other is no longer valid (the latter is coded as Questioned and the former as Followed) (Shepard's 1993: 21-22). Third, Limited signifies that the citing case has set a boundary on the precedential effect of the cited case by refusing to extend it beyond the exact issues involved or by stating that it only applies in specific, limited circumstances. Examples of Limited include, "We refuse to extend the rule of 'C v. D' to this situation," or "The rule in 'X v. Y' case shall be limited [or confined] to its own set of facts" (Shepard's 1993: 20). Finally, Criticized indicates the citing case disagreed with the cited case. It can be used, among other times, when: "the court in the citing case agrees with the dissent in the cited case," or "the discussion in the cited case is referred to as dicta and is rejected by the citing case" (Shepard's 1993: 14).

The remaining three treatments are not as "strong" as those above, meaning they have less impact on the precedential value of a cited case. Distinguished is a negative treatment used when the citing case states that it differs in a significant way from the cited case. Opinion language triggering this code can include: "... distinguishable, different, ... inapplicable, inapposit. . ." (Shepard's 1993: 15). The final two codes are considered as neither negative nor positive treatments of a cited case. Harmonized indicates that the citing and cited case differ in some way but the citing case reconciles the apparent inconsistency. Examples of Harmonized include: "We do not find the discussion of 'A v. B' to be inconsistent with the results reached herein"; or "this is a distinction without a difference" (Shepard's 1993: 18). *Shepard's* employs Explained when the citing case "clarifies, interprets, construes or otherwise annotates the decision in the cited case" (Shepard's 1993: 16). It would be used in language similar to the following: "Appellant cites case 'A v. B', as laying down the additional requirement that . . . However, the case of 'A v. B' does not hold that . . . but merely sets forth. . ." (Shepard's 1993: 16).

⁶ A court, as a matter of law, must have authority to overrule in order to warrant the use of this treatment code. If the citing court is without authority to overrule then *Shepard's* uses the Questioned code instead. Thus, for example, only a majority opinion can be coded as Overruling a cited case (Shepard's 1993: 20-21).

It is conceivable that the language of a citing opinion might suggest the use of two or more treatment categories. *Shepard's* developed two rules to determine which codes would be used in such a situation. First, *Shepard's* decided that multiple treatment codes could be assigned only if each one concerned a different point of law in the cited case (Shepard's 1993: 22-23). To determine points of law, Shepard's creates "headnotes" for each case, where each numbered headnote constitutes a separate point of law. Thus, for example, a citing case could both Follow headnote 1 and Distinguish headnote 2 of a cited case; a citing case, however, could not both Follow and Distinguish the same point of law in the cited case. Second, if two codes could be applied to the same point of law then Shepard's adheres to the "strongest letter applies" rule to determine which letter "predominates" (Shepard's 1993: 22). This rule arranges the treatment codes in terms of their strength. The order of strength is: Overruled, Questioned, Limited, Criticized, Followed, Distinguished, Explained, and Harmonized (Shepard's 1993: 24). Thus, if the citing opinion's language suggested that it both overruled and questioned the cited case then *Shepard's* codes the citing case as Overruling the cited case, since it is the stronger form of negative treatment.⁷

REPLICATING SHEPARD'S CITATIONS

Our replication proceeded in three steps. First, we determined whether *Shepard's* fails to list cases actually cited in court opinions. More specifically, we ascertained whether, for 25 randomly selected Supreme Court citing cases, *Shepard's* accidentally excluded any of the cited cases in them. Using the *United States Supreme Court Reports—Lawyer's Edition*, we located 300 cited cases in these 25 randomly selected opinions. We then "shepardized" each of these 300 cases to ascertain whether *Shepard's* listed it as being cited by the citing case.

Second, for each citing case *Shepard's* must determine whether the cited case is being legally interpreted by the citing opinion or is just being mentioned without any substantive legal treatment. For *Shepard's* to apply a treatment code, the citing case must contain language that legally treats the cited case. A mere string citation, for example, would not by itself warrant a treatment code, and such a case would be listed in *Shepard's* as citing, though not legally interpreting, the cited case. We must therefore ascertain whether *Shepard's* reliably determines when a citing case legally treats a cited case. To do so, we randomly selected 25

⁷ *Shepard's* relaxed the strongest-letter-applies rule in 1993, and starting late in that year a citing case can give multiple legal treatments to a cited case for the same point of law. John Strand, a long-time employee of the Shepard's Company, informed us that the strongest-letter-applies rule was originally issued to conserve on space in the printed product (phone conversations 8-6-98, 11-10-98, and 1-5-99). With the advent of electronic technology—and the dissemination of *Shepard's* via compact disk and on-line services such as Lexis—this concern no longer exists and thus Shepard's stopped using this rule in 1993.

U.S. Supreme Court cases decided between 1946 and 1987.⁸ We then used boolean word searches in Lexis for the litigant names and case citations to locate every subsequent Supreme Court opinion (through January 1999) referring to each of these 25 cases. This procedure resulted in a sample of 252 citing cases for these 25 cited cases (see footnote 12). We then read each citing opinion, coding it as either legally treating the cited case or just mentioning it without any substantive treatment.

Third, for each Supreme Court case decided between the 1946 and 1995 terms (a cited case), we used *Shepard's* to identify how all subsequent cases (citing cases) legally treated it. To replicate *Shepard's* coding of the citing cases' legal treatment of the cited cases, we took a 6 percent random stratified sample (N = 602) of all the citing cases in *Shepard's* for this time period. We used a stratified sample because it is the most appropriate way to ensure that all treatment categories are adequately represented in the reliability data (see Krippendorff 1980: 146). This sampling design is appropriate here since a number of the treatment categories do not occur very frequently.⁹ We then read all of the citing cases' majority and concurring opinions in the *U.S. Reports*, coding how each citing opinion legally treated the cited opinion using the coding rules located in the *Shepard's* (1993) in-house training manual.

RESULTS

Reliability

To determine if *Shepard's* is missing case cites, we examined the 300 cited cases located in the 25 randomly selected Court opinions. As one might expect (given the nonsubjective nature of this coding decision), this aspect of *Shepard's* data is exceptionally reliable. Indeed, we find that *Shepard's* had an accuracy rate of 100 percent in that it included each of the 300 cited cases as being referenced in the 25 cases. Thus, we conclude that *Shepard's* list of cited cases is most likely not underinclusive.¹⁰

⁸ We used 1987 as an end date to ensure that each cited case had enough time to be cited by the Court in subsequently decided cases.

⁹ While Overruled, Questioned, Criticized, Limited, and Harmonized respectively occur in only .6 percent, 2.5 percent, 1.3 percent, .2 percent, and 1.3 percent of citing cases' treatments of cited cases decided between the 1946-1995 terms, they respectively comprise 2.8 percent, 7.8 percent, 5.7 percent, 2.0 percent, and 4.7 percent of our reliability sample. The most frequent types of legal treatment a citing case gives a cited case in *Shepard's* data (for cited cases decided between the 1946-1995 terms) are: Followed (38.1 percent), Distinguished (31.2 percent), and Explained (24.9 percent). We did not conduct a reliability analysis of the Parallel code because *Shepard's* applied it only four times during this entire time period.

¹⁰ Songer (1988) notes two additional reasons *Shepard's* may not include a case that is on point. We do not think that either poses a problem for *Shepard's* data regarding the U.S. Supreme Court.

The second step in our reliability analysis was to ascertain whether, for a given precedent, *Shepard's* provides a reliable list of the subsequent cases that legally treat it. In addition to listing the cases that refer to a precedent, *Shepard's* determines how a citing case legally interprets the cited case. Thus, *Shepard's* must decide whether a case legally interprets, rather than just cites, a case. After coding the 252 citing Court cases that refer to the 25 randomly selected Court opinions, we find that our coding agrees with *Shepard's* 88.5 percent of the time. While this high rate of agreement indicates that the coding appears to be reliable, it is also necessary to consider the extent to which this agreement exceeds that expected by chance. The Kappa statistic performs such a calculation (see Cohen 1960).¹¹ In this case, the Kappa statistic for the agreement between our coding and *Shepard's* is .683 ($p < .001$).¹² This Kappa means that the level of agreement is 68.3 percent greater than would be expected by chance and thus indicates "substantial" intercoder agreement (see Landis and Koch 1977). Therefore, *Shepard's* designation of treatment cases appears to be reasonably reliable.

First, Songer argues that *Shepard's* only lists a case if the court opinion gives the full citation. According to *Shepard's* personnel (phone conversations with Leslie Martin 6-8-98, Rebecca Marshall 6-30-98, and John Strand 9-13-99), *Shepard's* includes any case that is referenced in an opinion, regardless of whether the full cite is given. If the full cite to a case is not given then *Shepard's* employees indicated to us that they make every effort to ascertain to which case the court opinion refers. In fact, they keep a list of popular names court opinions will sometimes use when referring to another case, such as "Erie Doctrine" or "Frye standard." Most importantly, *Shepard's* policy is to include references to cases, and they will not include a reference to a doctrine without a reference to the case. For example, if a court simply refers to the to the "Miranda rights," without referring to the case at some point in the opinion, then *Shepard's* will not list the case. Nevertheless, we do not anticipate this issue to pose a problem at the U.S. Supreme Court, because in our experience the Court almost always provides full citations to a cited case somewhere in the citing opinion. We, however, can not speak to the extent to which lower court opinions do not provide full citations. Second, Songer (1988) suggests that if a court defies a precedent, and thus does not refer to the precedent when it should, *Shepard's* will not list the case. Our data and analysis applies only to the U.S. Supreme Court, and we think it unlikely that the Court would systematically fail to cite a precedent when it should. Indeed, it would be difficult to even determine when the Court should cite a case, given the complexity and disagreement surrounding most Supreme Court cases. As for lower courts, recent research shows that lower court noncompliance with the Supreme Court is rare (see Songer, Segal, and Cameron 1994).

¹¹ $Kappa = (p_o - p_c) / (1 - p_c)$, where p_o is the observed proportion of agreement, and p_c is the proportion of agreement expected by chance (Cohen 1968). The Kappa statistic can range from below 0 to 1. If Kappa equals 0, then the amount of agreement between the two coders is exactly what one would expect by chance. If Kappa equals 1, then the coders agree perfectly. When evaluating the extent to which the two coders agree, Landis and Koch (1977) attach the following labels to the size of the Kappa statistic: < 0.00 is Poor; 0.00 – 0.20 is Slight; 0.21 – 0.40 is Fair; 0.41 – 0.60 is Moderate; 0.61 – 0.80 is Substantial; and 0.81 – 1.00 is Almost Perfect.

¹² While our sample is relatively small ($N = 252$), given the amount of time it takes to code these cases, the 95 percent confidence interval around the Kappa is a reasonably tight $\pm .121$.

The third step in our analysis was to assess the reliability of *Shepard's* assignment of specific treatment codes to the citing cases. First, we analyzed the reliability of the individual treatment codes and, for example, compared our assignment of the Followed code with that of *Shepard's*. The results of this analysis are presented in Table 1. Inter-coder agreement for these treatment codes varies from 79.1 percent to 99.5 percent, and the accompanying Kappa statistics range from .401 to .907, suggesting that all the treatment codes are at least moderately reliable.¹³ This said, the variation in agreement between the treatment codes is considerable. The treatment codes that *Shepard's* describes as negative (Distinguished, Criticized, Limited, Questioned, and Overruled) are the most reliable and have Kappa statistics ranging from .618 to .907. The stronger negative treatment codes, Limited, Questioned, and Overruled, are particularly reliable. While not quite as reliable as the negative treatments, the Followed code still exhibits substantial agreement (81.9 percent, Kappa = .608). The neutral treatment codes are the least reliable, as both Explained and Harmonized have Kappa statistics (.402 and .401, respectively) that are right on the cusp of "moderate" agreement.

Scholars may not always want to employ the individual treatment codes in their research. Instead, there may be occasions in which it is desirable to collapse the treatment codes into three broad categories: positive treatment, neutral treatment, and negative treatment (e.g., see Spriggs and Hansford 1998). For this reason, we also examined the reliability of assigning these directional treatments. Again, Followed is a positive treatment, Explained and Harmonized are neutral, and Distinguished, Criticized, Limited, Questioned, and Overruled collapse into the category of negative treatment. Based on the comparison of our coding with that of *Shepard's*, we find that the collapsed treatment categories are also quite reliable. The negative treatment code is the most reliable as the inter-coder agreement is 86.4 percent and the Kappa statistic is .725. The positive treatment category also falls within Landis and Koch's substantial agreement classification with a Kappa statistic of .608 (81.9 percent agreement). While not unreliable, the neutral treatment category is the least reliable of the three (78.2 percent agreement, Kappa = .450).¹⁴

Shepard's has slightly altered its coding rules over time in response to customer feedback and changes in information technology (see footnote 7). These changes could possibly cause the reliability of the data to vary over time. However, our analysis shows that the reliability of the data is quite stable over time. Specifically, we examined whether the rate of inter-coder agreement differed for

¹³ The confidence intervals around these Kappas are quite small, and the largest 95 percent confidence interval is $\pm .080$.

¹⁴ We believe that these differences in reliability result from variations in the specificity of *Shepard's* coding rules. While the coding rules for the negative treatment codes are highly specific and often require specific language to be present, the coding rules for the neutral codes are less precise and allow for more coder discretion.

≡ TABLE 1
 RELIABILITY OF SHEPARD'S CITATIONS' LEGAL TREATMENT CODES

	Intercoder Agreement (Kappa)		
	Entire Sample 1946-1995	Subsample 1 1946-1974	Subsample 2 1975-1995
<i>Individual Treatment Codes</i>			
Followed	81.9% (.608)	—	—
Explained	79.1% (.402)	—	—
Harmonized	94.7% (.401)	—	—
Distinguished	87.9% (.690)	—	—
Criticized	94.5% (.618)	—	—
Limited	99.0% (.781)	—	—
Questioned	96.4% (.713)	—	—
Overruled	99.5% (.907)	—	—
<i>Treatment Direction</i>			
Positive	81.9% (.608)	82.3% (.645)	81.7% (.593)
Neutral	78.2% (.450)	82.3% (.420)	76.7% (.451)
Negative	86.4% (.725)	84.8% (.692)	87.0% (.737)
N	602	164	438

Note: For all Kappas reported above, $p < .001$ that Kappa = 0.

citing cases decided in or after 1975 ($n = 438$), as compared to those decided before 1975 ($n = 164$).¹⁵ Since many of the individual treatment codes do not

¹⁵ We chose 1975 as a break point because the changes in the coding rules about which we are concerned occurred after 1975.

occur with much frequency in the pre-1975 sample, we focus attention on the broader treatment categories. As evident in Table 1, very little difference exists across these two time periods. For example, the proportion of agreement for positive treatment in pre-1975 citing cases was 82.3 percent (Kappa = .645), and it was 81.7 percent (Kappa = .593) for cases decided in or after 1975.

Validity

The above analysis demonstrates that *Shepard's* coding scheme is reasonably reliable, but are the categories valid representations of what they intend to capture? Unlike reliability, validity can not be assessed outside of a theoretical context that determines what is being measured (Carmines and Zeller 1979: 16). While we will not attempt to anticipate all possible uses of *Shepard's*, we will point out features of the data about which one should be aware when determining their validity for a particular purpose. Thus, the following discussion will point out certain characteristics of *Shepard's* coding scheme that can potentially affect validity, though the extent to which they actually do so will depend on the concepts being measured.

One of the limitations of *Shepard's* is that a few of the treatment categories are somewhat heterogeneous, meaning that qualitatively different types of legal treatments may be included within the same coding category. *Shepard's*, for example, assigns Distinguished for varying levels of negative treatment. It is often employed when the citing case merely states that there are a few, relatively minor differences between it and the cited case. In other instances, a citing case will clearly negatively interpret a cited case in the context of Distinguishing it. Thus, it is likely that some of the citing cases coded as Distinguished do not cast much doubt on the cited cases. In fact, the *Shepard's* manual indicates that Distinguished is the weakest form of negative treatment, and at least two *Shepard's* letter editors described it as relatively "unimportant" when compared to the stronger negative treatments.¹⁶

Other treatment categories appear heterogeneous as well. Limited, like Distinguished, can encompass different levels of negative treatment. It can be applied both when a citing case explicitly limits a cited case to a particular factual or legal situation and when the court refuses to extend the cited case to a new setting. While Questioned is applied when a citing case determines the cited case lacks continuing validity, the reason for the court drawing this conclusion can vary. Questioned can be applied when the citing case notes that either a prior court opinion or a statute overruled the cited case, as long as the citing case agrees with the overruling. Questioned can also result from a citing case strongly

¹⁶ Phone conversation on 7-7-98 with Janell Robinson and written correspondence (3-19-99) from Jane Morris, both *Shepard's* employees.

disparaging a cited case without any prior judicial or statutory changes. Even the Followed category can embrace at least two scenarios—the citing case applied the cited case to the same factual scenario as in the cited case or the citing case extended the precedential reach of the cited case. In this sense, one additional restriction of *Shepard's* is its lack of a category for the expansion of a precedent.

A second potential drawback of *Shepard's* is that the legal treatment denoted by Explained is ambiguous. In some respects, Explained appears to be a residual category that does not necessarily connote that a citing opinion legally interprets the cited opinion in any substantively meaningful way. Rather, it often indicates that the citing opinion explicitly discussed the cited opinion, but without either relying on it as precedent or in any way discrediting it. Thus, it is likely that many Explained treatments do not constitute substantive legal interpretations of a cited case.

RECOMMENDATIONS AND CONCLUSION

Our analysis indicates that *Shepard's* provides a reliable indicator of how citing cases legally treat cited cases. We are particularly sanguine about the reliability of the stronger negative treatment codes (Overruled, Questioned, Limited, and Criticized), while the neutral treatment codes (Harmonized and Explained) appear to be the least reliable. Thus, we believe that, at least when judged on a reliability basis, *Shepard's* represents an extremely valuable source for social scientists interested in law and legal change. Given our research design, of course, our results only explicitly apply to *Shepard's* U.S. Supreme Court data.

But, *Shepard's* is not a perfect data source and users should be aware of a few issues regarding the validity of its treatment codes. As we have discussed, some of the treatment categories are fairly heterogenous. For treatments such as Questioned and Limited, we suggest that, depending on the particular research question, scholars may want to consider reading the relevant text in the citing opinion so that it can be determined whether the treatment comports with the concept being measured. One might not want to include a Questioned code in a measure of the Court's negative treatment of a prior opinion, for instance, if it turns out that in the citing case the Court merely acknowledges that Congress has overridden the cited case. This judgement, of course, is left to the researcher and his or her measurement goals. Since the Limited and Questioned codes are relatively rare, this suggestion is not that onerous.

In some research contexts, *Shepard's* neutral treatment codes may not be the types of legal treatment in which social scientists will be particularly interested. The relative ambiguity of the neutral treatment codes can also create problems for a scholar attempting to build measures of legal change or treatment of previously decided cases. We suggest that for some purposes it might be best to exclude the neutral treatment categories from any such measures.

The fact that *Shepard's* also codes concurring opinions leads to one final rec-

ommendation. Most of *Shepard's* treatment categories can be assigned to concurring opinions (Followed and Overruled are explicitly reserved for majority opinions) and, depending on the nature of the research being pursued, this can lead to validity concerns. That is, a scholar may not want to include treatments occurring in concurring opinions if the principal interest is measuring the Court's legal treatment of a case. We think that researchers might consider consulting the text of the case to determine if the treatment in question originates from a concurring or majority opinion. Scholars can then determine if they want to include any treatments that occur in concurring opinions.¹⁷

To conclude, scholars interested in courts should devote more attention to studying court opinions, legal change, and the role of precedent. Judges and courts are ultimately important decision-makers because the legal rules undergirding their opinions can have political, social, and economic effects. To explain the development or change in law, we must therefore develop valid and reliable indicators of these concepts. One reasonable way to do so is by using *Shepard's Citations*. But, any such use must be done in an informed manner. This means that *Shepard's* should not be used indiscriminately without understanding the likely validity of the measures being constructed. With this recognition, *Shepard's* can provide a valuable data source for developing indicators of legal change, the treatment of precedent, and the like.

REFERENCES

- Baum, Lawrence. 1988. "Measuring Policy Change in the U.S. Supreme Court." *American Political Science Review* 82: 905-12.
- Benesh, Sara C., and Malia Reddick. 1998. "Overruled: An Event History Analysis of Lower Court Reaction to Supreme Court Alteration of Precedent." Presented at the annual meeting of the Midwest Political Science Association, Chicago.
- Brenner, Saul, and Marc Stier. 1996. "Retesting Segal and Spaeth's *Stare Decisis* Model." *American Journal of Political Science* 40 (November): 1036-48.
- Brenner, Saul, and Harold J. Spaeth. 1995. *Stare Indecisus: The Alteration of Precedent on the U.S. Supreme Court, 1946-1992*. Cambridge, MA: Cambridge University Press.
- Carmines, Edward G., and Richard A. Zeller. 1979. *Reliability and Validity Assessment*. Newbury Park, CA: Sage.
- Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1): 37-46.

¹⁷ It is relatively easy to determine whether a legal treatment is in a concurring opinion because *Shepard's* provides the page number on which the treatment occurs. *Shepard's* on-line service via Lexis now also provides a separate designation for whether the legal treatment occurs in a concurring opinion, but it does so only for cases decided in approximately the past ten years.

- Epstein, Lee, and Joseph F. Kobylka. 1992. *The Supreme Court and Legal Change: Abortion and the Death Penalty*. Chapel Hill: The University of North Carolina Press.
- Epstein, Lee, and Carol Mershon. 1996. "Measuring Political Preferences." *American Journal of Political Science* 40: 261-94.
- Gates, John B. 1990. "Content Analysis: Possibilities and Limits for Qualitative Data." *Judicature* 73 (December/January): 202-03.
- Gates, John B., and Glenn A. Phelps. 1996. "Intentionalism in Constitutional Opinions." *Political Research Quarterly* 48: 245-61.
- George, Tracy E., and Lee Epstein. 1992. "On the Nature of Supreme Court Decision Making." *American Political Science Review* 86: 323-37.
- Johnson, Charles A. 1986. "Follow-Up Citations in the U.S. Supreme Court." *Western Political Quarterly* 39: 538-47.
- Johnson, Charles A. 1987. "Content-Analytic Techniques and Judicial Research." *American Politics Quarterly* 15: 169-97.
- Kemper, Mark. 1998. "An Event-Count Model of Supreme Court Overruling Decisions, 1803-1993." Presented at the annual meeting of the Midwest Political Science Association, Chicago.
- Knight, Jack, and Lee Epstein. 1996. "The Norm of *Stare Decisis*." *American Journal of Political Science* 40 (November): 1018-35.
- Krippendorff, Klaus. 1980. *Content Analysis: An Introduction to its Methodology*. Beverly Hills, CA: Sage.
- Landis, Richard J., and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33: 159-74.
- Miranda v. Arizona*. 1966. 384 U.S. 436.
- Pritchett, C. Herman. 1948. *The Roosevelt Court: A Study in Judicial Politics and Values 1937-1947*. New York: MacMillan.
- Rohde, David W., and Harold J. Spaeth. 1976. *Supreme Court Decision Making*. San Francisco, CA: W.H. Freeman.
- Segal, Jeffrey, A. 1985. "Measuring Change on the Supreme Court: Examining Alternative Models." *American Journal of Political Science* 29: 461-79.
- Shepard's Company. 1993. "Shepard's Citations In-House Training Manual." Photocopy.
- Songer, Donald R. 1988. "Case Selection in Judicial Impact Research." *Western Political Quarterly* 41: 569-82.
- Songer, Donald R., and Stefanie A. Lindquist. 1996. "Not the Whole Story: The Impact of Justices' Values on Supreme Court Decision-Making." *American Journal of Political Science* 40(November): 1049-63.
- Songer, Donald R., Jeffrey A. Segal, and Charles M. Cameron. 1994. "The Hierarchy of Justice: Testing a Principal-Agent Model of Supreme Court-Circuit Court Interactions." *American Journal of Political Science* 38: 673-96.

- Spaeth, Harold, and Jeffrey A. Segal. 1999. *Majority Rule or Minority Will: Adherence to Precedent on the U.S. Supreme Court*. New York: Cambridge University Press.
- Spriggs, James F., II. 1996. "The Supreme Court and Federal Administrative Agencies: A Resource-Based Theory and Analysis of Judicial Impact." *American Journal of Political Science* 40: 1122-51.
- Spriggs, James F., II., and Thomas G. Hansford. N.d. "Explaining the Overruling of U.S. Supreme Court Precedent." *Journal of Politics*. Forthcoming.
- Wahlbeck, Paul J. 1997. "The Life of the Law: Judicial Politics and Legal Change." *Journal of Politics* 59(August): 778-802.
- _____. 1998. "The Development of a Legal Rule: The Federal Common Law of Public Nuisance." *Law & Society Review* 33 (3): 613-68.

Received: September 20, 1999
Accepted for Publication January 31, 2000
jfspriggs@ucdavis.edu
tghansford@ucdavis.edu